A BAYESIAN LONGITUDINAL CIRCULAR MODEL AND MODEL SELECTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ONUR ÇAMLI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

AUGUST 2017

Approval of the thesis:

A BAYESIAN LONGITUDINAL CIRCULAR MODEL AND MODEL SELECTION

submitted by **ONUR ÇAMLI** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. Ayşen Dener Akkaya
Head of Department, **Statistics** _______________

Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu
Supervisor, **Department of Statistics, METU** _______________

Examining Committee Members:

Assoc. Prof. Dr. Çağdaş Hakan Aladağ
Department of Statistics, Hacettepe University _______________

Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu
Department of Statistics, METU _______________

Assoc. Prof. Dr. Kasırga Yıldırak
Department of Actuarial Sciences, Hacettepe University _______________

Assoc. Prof. Dr. Özlem İlk Dağ
Department of Statistics, METU _______________

Assist. Prof. Dr. Muhammet Burak Kılıç
Department of Business, Mehmet Akif Ersoy University _______________

Date: _______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ONUR ÇAMLI

Signature :

ABSTRACT

A BAYESIAN LONGITUDINAL CIRCULAR MODEL AND MODEL SELECTION

ÇAMLI, ONUR
M.S., Department of Statistics
Supervisor : Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu

August 2017, 88 pages

This research was motivated by a recent medical study that aims to estimate the general fetal head progression trajectory during the first stage of normal labour adjusted for maternal characteristics and environmental factors. A rather primitive manual method for determining the progression has recently been replaced by an ultrasound technology that can precisely measure the fetal's head angle. The particular challenge with such data is the model selection procedures that could objectively assess the models when outcome data are longitudinal and circular. A Bayesian random intercept model on the circle was considered and the current model selection methods used in Bayesian analysis of circular data were reviewed and commented. Then criteria based on minimizing a predictive loss was focused and some new methods and new extensions to a current method were proposed. Extensive Monte Carlo simulation studies controlled for the sample size and intraclass correlation were used to study the performances of the model and these model selection criteria under various realistic longitudinal settings. Relative bias and mean square error were used to evaluate the performance of the estimators under correctly specified models and robustness to model misspecification. Several quantities were used to evaluate the performances of the model selection criteria such as frequency of selecting the true model and a ratio that measures the strength of the particular selection. Simulations reveal a noticeable or equivalent gain in performance achieved by the proposed methods. A conventional longitudinal data set (sandhopper data) was used to further compare the Bayesian model selection methods for circular data. This research hopes to address and contribute to the model selection in circular data, a rather fertile area for methodological and theoretical development, while the demand increases with the advancing technology as seen in our motivating data set.

Keywords: Directional Statistics, Random Effects, Model Selection, Medicine, Biology.

ÖZ

BAYESCİ UZUNLAMASINA DAİRESEL BİR MODEL VE MODEL SEÇİMİ

ÇAMLI, ONUR
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi : Doç. Dr. Zeynep Işıl Kalaylıoğlu

Ağustos 2017 , 88 sayfa

Normal doğum eyleminin ilk evresinde anneliğe özgü özelliklere ve çevresel faktörlere göre şekillenen genel cenin baş ilerleme yörüngesini tahmin etmeyi amaçlayan yakın tarihli bir tıbbi araştırma gerçekleştirilen tez çalışmasına motivasyon kaynağı olmuştur. Daha önceleri ilerlemeyi belirlemek için oldukça ilkel elle yapılan bir yöntem günümüzde yerini ceninin baş açısını tam olarak ölçebilen bir ultrason teknolojisine bırakmıştır. Böyle veriler ile ilgili zorluk, sonuç verisi uzunlamasına ve yönsel veri olduğunda modelleri tarafsızca değerlendirebilen model seçim yöntemleridir. Çember üzerinde Bayesci rassal kesen içeren bir model düşünülmüş ve açısal verilerin Bayesci analizinde kullanılan güncel model seçme metodları gözden geçirilerek değerlendirilmiştir. Daha sonra tahmini bir kayıbın minimum düzeye indirilmesine dayalı olan kriterlere odaklanılmıştır. İki tane yeni metot ve var olan bir yöntem için yeni bir genişletme önerilmiştir. Çeşitli gerçekçi uzunlamasına senaryolar altında modelin ve model seçim kriterlerinin performanslarını incelemek için, örneklem hacmi ve sınıf içi korelasyonu için kontrol edilen kapsamlı Monte Carlo simülasyon çalışmaları yapılmıştır. Modelin doğru belirlenmesi durumunda tahmin edicilerin performanslarını ve yanlış model belirlenmesine karşı sağlamlıklarını değerlendirmek için nisbi sapma ve hata kareler ortalaması ölçüleri kullanılmıştır. Model seçme kriterlerinin performanslarını değerlendirmek için doğru modeli seçme sıklığı ve belli bir seçimdeki kararlılığı ölçen bir oran gibi birkaç ölçüt kullanılmıştır. Simülasyonlar, önerilen yöntemlerle elde edilen performansların dikkat çekici veya eşdeğer olduğunu ortaya

koymuştur. Açısal veriler için Bayesci model seçme kriterlerini daha fazla kıyaslamak için bilinen uzunlamasına bir veri seti (kum çekirgesi veri seti) kullanılmıştır. Bu araştırma, motive edici veri setimizde görüldüğü gibi gelişen teknoloji ile talep arttıkça, metodolojik ve teorik gelişme için oldukça verimli bir alan olan yönsel verilerde model seçimine değinmeyi ve katkıda bulunmayı ummaktadır.

Anahtar Kelimeler: Yönsel İstatistikler, Rassal Etkiler, Model Seçme, Tıp, Biyoloji.

To my family

ACKNOWLEDGMENTS

I am very grateful to my thesis advisor Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu for her invaluable support, guidance, help, and contribution not only to this thesis but also to my real life. She has contributed to me in many ways especially to my academical view and vision of world significantly. Whenever I needed her great experiences, she patiently shared with me her worthful experiences. She also answered all my questions and directed me to reach my destination. She always encouraged me to work more. I have gained valuable experience while studying with her. It would have been impossible for me to finish this thesis without her direction, effort, and encouragement. It was an immeasurable honor and pleasure for me to work with her throughout the research.

I would like to thank my committee members, Assoc. Prof. Dr. Çağdaş Hakan Aladağ, Assoc. Prof. Dr. Zeynep Işıl Kalaylıoğlu, Assoc. Prof. Dr. Kasırga Yıldırak, Assoc. Prof. Dr. Özlem İlk Dağ and Assist. Prof. Dr. Muhammet Burak Kılıç for their time in detailed reviewing my thesis, valuable suggestions and comments.

I owe special thanks to my housemate Serdar Halis, my friends Adem Kına and Melih Ağraz their motivation, encouragement and nice friendship. I would also like to thank my officemate Elif Akça for her endless support, help and kind friendship. I am also grateful to Dr. M. Talha Arslan for his encouragement and invaluable support. I would also like to thank all the members of the Department of Statistics.

Finally, I would like to express my deepest love, warm thanks to my parents Hafize and Necmettin, to my sister Şeyma and to my brother Ömer for their complimentary love, smiley face and encouragement to reach my goals throughout my life. They have been always with me whenever I needed their support.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
APE	Average prediction error
APE_1	Absolute Predicted Errors-Type 1
APE_2	Absolute Predicted Errors-Type 2
ASC	Average Stability Coefficient
BIC	Bayesian Information Criterion
BGR	Brooks-Gelman-Rubin
CPD	Circular Predictive Discrepancy
CPD_1	Circular Predictive Discrepancy-Type 1
CPD_2	Circular Predictive Discrepancy-Type 2
CPO	Conditional predictive ordinate
DIC	Deviance Information Criterion
E	Expectation
EM	Expectation–Maximization
ES	Expectation solution
Eye	Eye Symmetry Index
GAIC	Generalized Information Akaike Criterion
ICC	Intraclass correlation coefficient
LCRIM	Longitudinal Circular Random Intercept Model
LPML	Log pseudo-marginal likelihood
M	Maximization
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MSE	Mean Square Error
MVM	Multivariate von Mises
$PAPE_1$	Plug-in Absolute Predicted Errors-Type 1
$PAPE_2$	Plug-in Absolute Predicted Errors-Type 2
pdf	Probability density function

PN	Projected Normal
RB	Relative Bias
TED	Theoretical Escape Direction
TM	True Model
vM	Von Mises
WC	Wrapped Cauchy
WN	Wrapped normal

CHAPTER 1

INTRODUCTION

1.1 A Motivating Dataset and Overview

Our research is motivated by an ongoing observational study at Ankara University, Department of Gynecology. The aim of the study is to develop a model for the fetus's head progression in the first stage of labour for normal birth based on the angle of head shown in Fig 1.1.

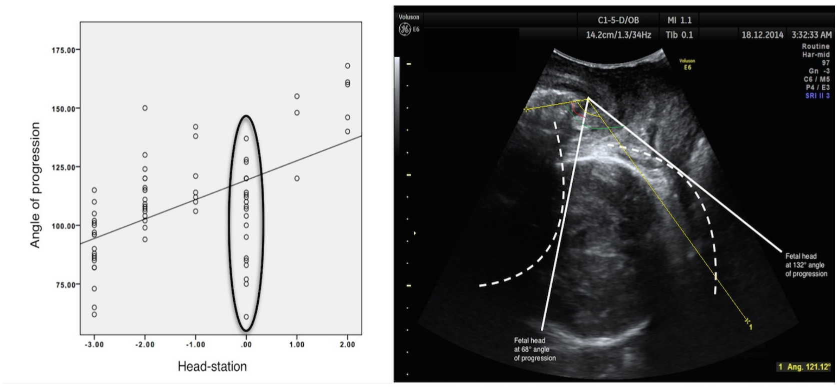

Figure 1.1: (Left) Regression plot showing relation between palpated head stations and angle of progression. (Right) Illustration of the limits of actual fetal head placements according to angle of progression with ultrasonography (Yuce et. al., 2015)

Head progression shows the movement of the fetus towards the outer world. Gynecologists use such models as reference curves to determine whether the head progression of the fetus in a particular first stage labour is normal so that the birth will result in a normal birth. The reference curve that is currently in use was constructed based on cervical dilation (Zhang et. al., 2010). However there is a new technique based on an ultrasound technology that measures the head progression angle precisely. The aim of the gynecological researchers at Ankara University is to construct a new reference curve using the angle data obtained by this technology and adjust it for maternal characteristics. Using this reference line, the gynecologist will be able to better tell whether a particular fetus's head angle is compatible with the angles observed in a situation that resulted in a normal birth. The dataset consists of angle data representing the movement of the head of the fetus and maternal characteristics. Angles are recorded about every 30 minutes during the first stage of the labour. Maternal characteristics include baseline characteristics and time dependent characteristics measured at the time of head angle recording such as current age, weight and height, parity, abortion history, and cervical dilation. The response data obtained in this study has two properties; circular and repeated and it is called longitudinal circular data. Since motivating data set is not complete yet, it will not be pursued further in the application section.

Repeated and circular observations (i.e. longitudinal circular data) arise in many different researches in fields such as medicine, meteorology, biology, geology and psychology. For instance, animal scientists wanted to understand the factors effecting the direction the sandhoppers jump to and so they recorded the direction each time a sandhopper moved (Scapini, 1997). All such examples require the use of circular statistics for longitudinal circular data. There have been a lot of statistical methods to analyze longitudinal data in linear structure. For example, see Diggle et al. (2002), Fitzmaurice et al. (2004), Hedeker and Gibbons (2006), Gelman and Hill (2007), and Fitzmaurice et al. (2008). However, the statistical analysis of circular longitudinal response data is a fertile area of statistics. There is a limited number of literature on statistical modeling of longitudinal circular response data. Also, model assessment/comparison/selection are not well addressed in circular data literature, particularly for complex modeling such as longitudinal modeling. Longitudinal structures and the difficulties in working with circular distributions may play an important role in these limitations. Our focus in this thesis is modeling and model selection for circular longitudinal responses.

1.2 Literature Review

There are several methods to analyze longitudinal response data on $\mathbb{R}$. However, methods for the analysis of longitudinal circular response data, i.e. data on S, is still limited in the literature. Current frequentist and Bayesian literature for longitudinal circular regression are given below.

Literature Review on Longitudinal Circular Regression

Artes et al. (2000) considered estimating equations for longitudinal circular data. They proposed a method based on the quasi-likelihood theory of Wedderburn (1974) and the generalized estimation equations approach of Liang and Zeger (1986). Their estimating function was defined as a measurable function of the data and the parameters of interest. They derived estimating equations for modeling the mean resultant length and the circular mean separately. They then derived estimating equations for a mixed model and examined the consistency and asymptotically normality properties of the estimators. They used the Newton method, the Newton modified method and one based on the modified Newton method.

D'Elia (2001) developed a variance components model for longitudinal circular data in which response variable model includes both fixed effects of explanatory variables and random components which consists of random intercept and temporal stochastic process. Response variable is assumed to have a von Mises distribution. They considered at least two different sources of random variation, between subject variation and within subject correlation. These are incorporated into the model framework through the random intercept and the temporal stochastic process respectively. Random intercept is assumed to have a normal distribution with zero mean and variance. Temporal stochastic process is assumed to have Stationary Gaussian Process. They used a simulated maximum likelihood approach to estimate the model parameters.

McMillan et. al. (2013) also used random intercepts to account for the correlation between the repeated circular measurements proposed for a two-part mixture of wrapped Cauchy densities. Proposed model can be used for modeling repeated measurements of bimodal angular data. This model is based on the work of McMillan et. al. (2011) in which the maximum likelihood is used for inference on the two-part wrapped Cauchy model.

Hall et. al. (2015) considered an approach to fit a marginal regression model for these data sets. This approach is a marginal version of the spherically projected multivariate linear model and it makes it possible the use of a working correlation matrix and a robust variance-covariance estimator. Hence, someone who want to use this approach should determine a working correlation matrix. They proposed this model to fit using an extension of the expectation–maximization (EM) algorithm in which the maximization (M) step is replaced by the solution of a generalized estimating equations. As a matter of principle, the EM algorithm can be applied to fit this model, but there are some practical obstacles. One of the obstacles is that there is the challenge of correctly modeling the correlation matrices and the variance of cluster- specific latent response. Another obstacle is that to calculate the expectation (E) step we need to calculate some terms which have no closed form, therefore an intractable multidimensional integral must be evaluated. To overcome these obstacles, they formulate the EM algorithm under independence, by introducing a working correlation matrix into the estimating equations and this methodology is known as expectation solution (ES) algorithm.

Lagona (2016) developed a flexible model based on the multivariate von Mises distribution (MVM). Both mean and concentration parameters of this distribution depend on some covariates through some link functions. In that study, a regression model that can accommodate various correlation structures such as heteroscedasticity, unstructured correlation, and specific autoregressive correlation structures is suggested. This proposed model can be considered as a multivariate extension of the generalized linear model developed by Fisher and Lee (1992) for independent circular response. The maximum likelihood estimation technique is used to obtain the model parameters by maximizing the log-likelihood function of model. However, this estimation method arises some problems. When the bivariate case is concerned, since the normalizing constant of MVM density can be derived, the estimates of parameters can be obtained. On the other hand, when the dimension is larger than two, the analytical form of the normalizing constant of MVM density cannot be obtained and numerical integration is not applicable. Therefore, to obtain the estimation of model parameters, a Monte Carlo approximation of the log-likelihood is used. To maximize the Monte Carlo approximation of the log-likelihood, they used classic Newton type procedures by using the properties of centered MVM distribution. The MVM density is not closed under marginalization and its normalizing constant is challenging, for this reason they used a simple Gibbs sampling to obtain samples from the MVM density. These distributional properties of the MVM distribution still make difficult the estimation of the model parameters.

Maruotti et. al. (2016) used the mixed projected normal distribution to deal with longitudinal circular data. They introduced a time-dependent mixed effect projected normal regression model based on a hidden Markov heterogeneity structure. Their model can be considered as an extension of the model developed by Presnell et. al (1998), since the conditional distribution of the circular response is projected normal. The developed model is useful tool for modeling multivariate circular variables and quite general. They used a maximum likelihood approach based on the radial projection of a bivariate normal distribution to make inference on model. An expectation–maximization algorithm is used to obtain the maximum likelihood estimation of the model parameters. However, their estimation produce does not produce standard errors of the estimates, since approximations based on the observed information matrix often require very large sample size. To get through this problem, they used parametric bootstrap approach and then calculated the approximate standard error of each parameter estimator.

Under the longitudinal and circular structure, Maruotti (2016) considered the implementation of flexible circular regression model, thus, the projected normal regression model is extended in two directions. To analyze the longitudinal circular data, first of all they used correlated random coefficients to introduce dependence between projections, and then defined a mixed effects model for radial projections onto the circle. Finally, they extended the variance component model methodology. The method can be considered an extension of the method developed by Presnell et. al. (1998). However, the proposed model is a semi-parametric extension of the projected normal model, because there is no any assumption on the distribution of random effects and any limitation on the covariance structure of random effects. To obtain the maximum likelihood estimations of the proposed semi-parametric model parameters, the EM algorithm was used again without any parametric assumption on the random effects distribution. At first appearance, the proposed model may be seen very attractive, but there are some computationally disadvantages and restrictions. In addition, their parameter estimation procedure does not generate standard errors of the estimates. Therefore, to obtain standard errors of the parameters, a parametric bootstrap technique is employed.

Rueda et. al. (2016) developed a flexible piecewise circular regression model for analysis of longitudinal circular data. In their model, both response variable and covariate variable are circular and this model works under the assumption that these circular variables have the von Mises distribution which is the most important and key distribution for circular data. Although the proposed model can allow for flexible relationships between covariate and response variable in different sectors (subjects or units), it can be considered as an extension of the model introduced by Downs and Mardia (2002). To make inference on their model, they used the maximum likelihood estimation method. However, there are some constraints on the parameters for existence of continuous solution. To obtain the maximum likelihood estimates for their model, they rewrite the model as the piecewise circular-linear model, and then they describe the derivation of maximum likelihood estimators by using the theory developed by Fisher and Lee (1992) for circular–linear model and the R package of Agostinelli and Lund (2011). The drawback of their model is that the estimation of the model parameters can be obtained when there is only one circular regressor variable since they used the circular–linear model developed by Fisher and Lee (1992) in their estimation procedure.

In Bayesian analysis framework, Antonio et. al (2014) developed a circular longitudinal model to analyze short series of longitudinal circular data. Their proposed model is based on projected normal distribution. In their model, a mixed effect linear model is used to determine each of the two components. They used a Metropolis within Gibbs scheme to obtain inferences about the model parameters.

Literature Review on Model Selection for Longitudinal Circular Regression

Selection of an appropriate approximating model is critical to statistical inference. As mentioned above, when a circular data is given, there is some class of parametric and semi-parametric models that can be fitted. In literature, there are a lot of model selection criteria used to select, compare and assess models with linear variables, but when modeling with circular variables is of concern, these usual model selection criteria employed for linear data may not be appropriate, like the other usual statistical tools not being suitable for analysis of circular data; as they do not account for the circular nature of the data. When modeling with circular variables is of concern, D'Elia (2001) used a test statistic which is based on the likelihood ratio criterion to compare some nested models and then by using the Akaike Information Criterion (AIC) (Akaike, 1973), they preferred the model which has fewer parameters. Carnicero et. al. (2008) used the Bayesian Information Criterion (BIC) (Schwarz, 1978), to select the number of terms to include in their mixture model. McMillan et al. (2013) employed the log pseudo-marginal likelihood (LPML) of Geisser and Eddy (1979) as a criterion for model selection and they preferred the model which maximizes the LPML. Rueda et. al. (2016) used the Generalized Information Akaike Criterion (GAIC) to assess the performance of models and they preferred the model that has the largest GAIC.

All these model selection criteria based on information theory and they were not developed especially for the circular nature of the data. Recently few model selection criteria specifically developed for circular data emerged. A model selection criterion which will be detailed in Chapter 4 was developed by Ravindran et. al. (2011) by using the decision theoretic framework of Gelfand and Ghosh (1998). This model selection criterion is called Circular Predictive Discrepancy (CPD). Maruotti et. al. (2016) used the circular distance suggested by Jammalamadaka and SenGupta (2001) to define the average prediction error (APE) to compare the prediction abilities of the models. They also used the AIC and the BIC to select the number of hidden states in their model which based on a hidden Markov heterogeneity structure. Maruotti (2016) again used the penalized likelihood criteria, as the AIC and the BIC, to determine the number of mixture components in their model which is based on finite mixture models. They also used the APE to compare the performance of models.

1.3 Our Objectives and Scope of the Study

As presented in Section 1.2, there are currently two basic approaches to regression modeling of longitudinal circular response data. These are i. use of multivariate circular distributions (estimating equations), ii. random effects models using projected normal distribution. The main drawback with these approaches are, respectively i. high dimensionality problem with multivariate distributions when within sample size (number of repeated measurements (i.e. longitudinal observations)) is large, ii. projected normal approach doubles the dimension of latent random effects space. We consider a modeling strategy that avoids these problems.

Another challenge with such data is the model assessment, comparison and selection procedures that objectively assess the models when outcome data are longitudinal and circular/directional. In this thesis, we focus on predictive model assessment. Performances of traditional criteria employed for longitudinal linear predictive model assessment such as AIC, BIC, DIC versions (Celeux et al., 2006), and prediction error based statistics are largely unknown for circular variables. We focus on the prediction error based model assessment methods.

The aims of this thesis are then listed as follows;

- Develop a new class of Bayesian models for longitudinal circular responses based on random intercept where response variable is circular and covariates are linear.
- Develop some novel predictive model assessment tools that take the account of circular properties of the data.
- Evaluate and compare the performances of model selection approaches.
- Illustrate the utilization of the considered model and new model selection criteria by using a real longitudinal angular data set.

This thesis is organized as follows. Chapter 2 gives some preliminaries and explains some basic definitions for circular data. Then, some properties of the most common circular distributions, such as the von Mises, the wrapped Cauchy and the projected normal distributions are presented in this chapter. The theory of linear circular regression model is also presented in this chapter. Chapter 3 gives some basic information about longitudinal circular data and explains the Bayesian methodology of with random intercept model, Bayesian analysis of the model which contains getting full conditional distributions, sampling methods, parameter estimations and implementation of the Markov Chain Monte Carlo (MCMC) scheme. Chapter 4 gives the proposed model selection criteria which are based upon predictive loss. Some properties of these model selection criteria and the selected model selection criteria currently used in practice are presented in this chapter. Chapter 5 gives the details of the extensive simulation study illustrating the performance of our considered model, our estimation procedure and comparing the performances of proposed model selection criteria and the other criteria. Application to a real data set is also presented in Chapter 6. Chapter 7 concludes the study with some discussion.

CHAPTER 2

PRELIMINARIES OF CIRCULAR DATA ANALYSIS

This chapter is concerned with the basic properties of circular measurements, basic summary statistics, most common circular distributions such as the von Mises distribution and the wrapped Cauchy distribution, the simplest theory of linear- circular regression model in which the relationship between a circular response variable and some linear covariates can be investigated. Researcher can come across circular data in a number of different areas such as meteorology, medicine, biology, earth sciences, physics, geology and psychology. and hence they should learn to analyze circular data as well as to assess and to compare circular models. For instance, in meteorology, wind directions or the times of year at which heavy rain occurs are circular data (Mardia and Jupp, 2000). In biology, the orientation of animals such as the direction of homing pigeons and escaping of the sand hoppers are circular data. In medicine the positional changes of the fetus in the birth canal, the fetal head progression in the first stage of labor are circular data. Data that are represented on unit circle, such as angles, are called circular data. For instance, directions, measurements recorded over a 24 hour period or a week or a year, are also circular. Circular data can be measured in degree or radian with the range ($0°$, $360°$) or (0, 2π) respectively. The starting points and endpoints are the same location on the circle for circular data. Since the angular value depends on the choice of the initial direction and the sense of rotation; clockwise or counter-clockwise, the numerical representation of a circular observation is not unique. Anyone who takes East as the initial direction and anticlockwise as the direction of rotation can consider the value of a circular observation as $75°$ and the value of the same circular observation can be considered as $15°$ by someone else who takes North as the initial direction and clockwise as the direction of rotation.

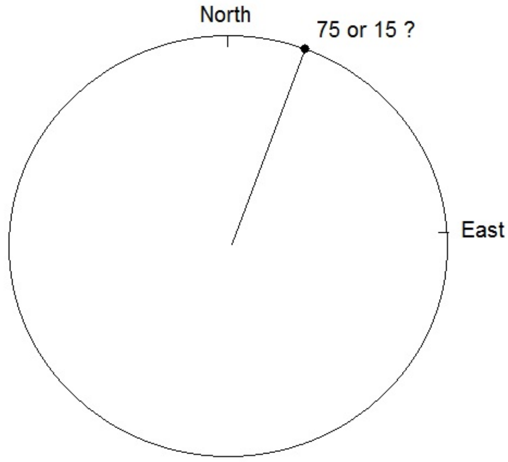

Figure 2.1: Varying value of a circular observation.

From Figure 2.1, it can be easily concluded that the choice of initial direction and rotation effect the representation of a circular observation. As mentioned above, circular data can be represented as angles or as points on the circumference of a unit circle and may be uniquely determined by two coordinates. The rectangular (Cartesian) coordinate system with origin 0 and polar coordinate system can be used for this purpose. Any point on the plane can be represented by using its rectangular coordinates or its polar coordinates as (X, Y) or (r, θ), respectively. (X, Y) are two perpendicular axes through the origin. r is the radius, θ is its direction. When a point in Cartesian Coordinates (X, Y) is given then by using the trigonometric functions sine and cosine, its counterpart in Polar Coordinates (r, θ) can be obtained. From Figure 2.2 the rectangular coordinates of a point are given in terms of its polar coordinates by

$$x = r\cos(\theta), y = r\sin(\theta). \tag{2.1}$$

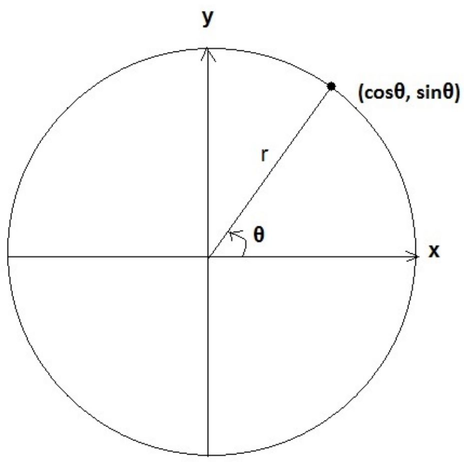

Figure 2.2: Rectangular and polar coordinates for an circular observation.

2.1 Basic Summary Statistics

Standard statistical approaches, such as descriptive statistics and classical modeling are not suitable for circular data and differ from those of linear methods. For example, let the directional data be x = (12°, 348°). The simple mean is 180° which is obviously incorrect as it is the opposite direction that does not represent the average data. The true simple mean is 0°. Because of such problems, an exclusive set of statistical approaches have been developed for analyzing circular data including descriptive statistics, statistical distributions and modeling (Fisher, 1993; Mardia and Jupp, 2000; Jammalamadaka and SenGupta, 2001).

2.1.1 Mean Direction and Mean Resultant Length

Let $\theta_1, \ . \ . \ . \ , \theta_n$ be a random sample of circular observations of size n on the unit circle. From Figure 2.2, let $x_i = \cos(\theta_i)$, $y_i = \sin(\theta_i)$ be polar transformation for each observation. Then, define $\bar{x} = \frac{1}{n}(x_1 + ... + x_n) = \frac{1}{n}(\cos(\theta_1) + ... + \cos(\theta_n))$ and $\bar{y} = \frac{1}{n}(y_1 + ... + y_n) = \frac{1}{n}(\sin(\theta_1) + ... + \sin(\theta_n))$. Then mean resultant length $\bar{R}$ is given by

$$\bar{R} = (\bar{x}^2 + \bar{y}^2)^{1/2}. \tag{2.2}$$

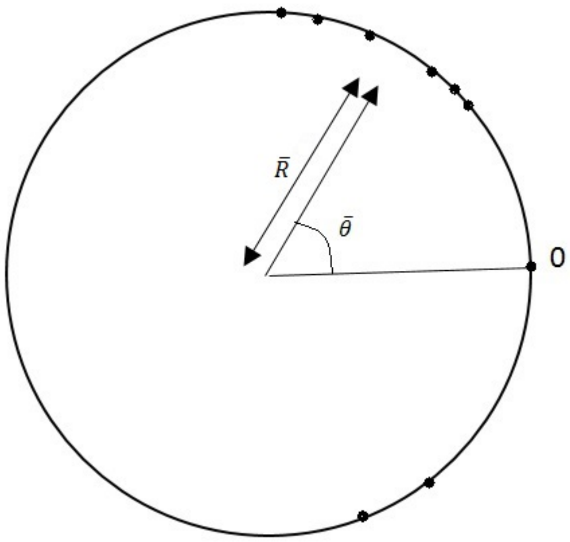

Figure 2.3: The mean resultant length $\bar{R}$ and the angle of mean direction $\bar{\theta}$ for a set of circular observation (Mardia, 1792)

The angle of mean direction is given by

$$\bar{\theta} = \begin{cases} \arctan(\frac{\bar{y}}{\bar{x}}) & \text{if } \bar{x} > 0, \bar{y} \geqslant 0, \\ \arctan(\frac{\bar{y}}{\bar{x}}) + 2\pi & \text{if } \bar{x} \geqslant 0, \bar{y} < 0, \\ \arctan(\frac{\bar{y}}{\bar{x}}) + \pi & \text{if } \bar{x} < 0, \\ \frac{\pi}{2} & \text{if } \bar{x} = 0, \bar{y} > 0, \\ undefined & \text{if } \bar{x} = 0, \bar{y} = 0. \end{cases} \tag{2.3}$$

The relationship between the mean resultant length and mean direction is

$$\cos(\bar{\theta}) = \frac{\bar{x}^2}{\bar{R}}, \quad \sin(\bar{\theta}) = \frac{\bar{y}^2}{\bar{R}}. \tag{2.4}$$

From (2.3) and (2.4), it can be concluded that the angle of mean direction is not equal $\frac{1}{n}(\theta_1 + ... + \theta_n)$.

2.1.2 Circular Variance and Circular Standard Deviation

The resultant length of the resultant vector is defined by

$$R = \bar{R}n \tag{2.5}$$

and lies in the range (0, n) where $\bar{R}$ is the mean resultant length which was given in 2.2 and lies in the range (0, 1). For circular data, when $\bar{R}$ is close to 1, this implies that the data set is highly concentrated and when $\bar{R}$ is close to 0, the data set is the less concentrated. $\bar{R}$ sometimes can not be suitable to show the dispersion of circular data. Then, the circular variance is defined by

$$V = 1 - \bar{R} \tag{2.6}$$

and lies in the range (0,1). The variation in the circular observation about the mean direction can be measured by using the circular variance. As the value of the mean resultant length decreases, the circular variance increases. However, when the circular variance is equal to 1, this does not imply that the data set has maximum dispersion.

The circular standard deviation is defined by

$$v = (-2\log(1 - V))^{1/2}. \tag{2.7}$$

2.1.3 Circular Symmetry Coefficient

The circular symmetry coefficient (s) is defined (Fisher, 1993) as follows

$$s = \frac{\rho_2 \sin(\mu_2 - 2\bar{\theta})}{(1 - \bar{R})^{\frac{3}{2}}}, \text{where } \rho_2 = \frac{1}{n}\sum_{i=1}^{n} \cos 2(\theta_i - \bar{\theta}) \tag{2.8}$$

and

$$\mu_2 = \begin{cases} \tan^{-1}(\bar{S_2}/\bar{C_2}) & \text{if } \bar{S_2} > 0, \bar{C_2} > 0, \\ \tan^{-1}(\bar{S_2}/\bar{C_2}) + \pi & \text{if } \bar{C_2} < 0, \\ \tan^{-1}(\bar{S_2}/\bar{C_2}) + 2\pi & \text{if } \bar{S_2} < 0, \bar{C_2} > 0, \end{cases} \tag{2.9}$$

where,

$$\bar{S_2} = \frac{1}{n}\sum_{i=1}^{n} \sin(2\theta_i),\ \bar{C_2} = \frac{1}{n}\sum_{i=1}^{n} \cos(2\theta_i). \tag{2.10}$$

Here, $\bar{\theta}$ is the mean direction defined in equation 2.3 and $\bar{R}$ denotes the mean resultant length defined in equation 2.2. s is nearly zero for symmetric unimodal data sets.

2.1.4 Circular-Circular Association

Let (Φ_1, Ψ_1) and (Φ_2, Ψ_2) be two independent random vectors of (Φ, Ψ). The circular correlation coefficient is defined by (Fisher and Lee, 1983)

$$\rho_T = \frac{E[\sin(\Phi_1 - \Phi_2)\sin(\Psi_1 - \Psi_2)]}{\{E[\sin^2(\Phi_1 - \Phi_2)]E[\sin^2(\Psi_1 - \Psi_2)]\}^{1/2}}. \tag{2.11}$$

This quantity explains association between two circular random variables. It takes values between -1 and 1. If Φ and Ψ are independent, then $\rho_T = 0$.

Given a random sample of n observations of (Φ, Ψ), (ϕ_1,ψ_1), ... , (ϕ_n,ψ_n), the estimate of $\hat{\rho_T}$ is calculated as follows

$$\hat{\rho_T} = \frac{4(AB - CD)}{\{(n^2 - E^2 - F^2)(n^2 - G^2 - H^2)\}^{1/2}} \tag{2.12}$$

where,

$A = \sum_{j=1}^{n} \cos\phi_j \cos\psi_j$, $B = \sum_{j=1}^{n} \sin\phi_j \sin\psi_j$,

$C = \sum_{j=1}^{n} \cos\phi_j \sin\psi_j$, $D = \sum_{j=1}^{n} \sin\phi_j \cos\psi_j$,

$E = \sum_{j=1}^{n} \cos 2\phi_j$, $F = \sum_{j=1}^{n} \sin 2\phi_j$,

$G = \sum_{j=1}^{n} \cos 2\psi_j$, $H = \sum_{j=1}^{n} \sin 2\psi_j$.

2.1.5 Circular Distance

Let x_1 and x_2 be any two points on circle and θ_1 and θ_2 be two angles corresponding to x_1, x_2 respectively. The circular distance between these two points can be measured by using

$$d(\theta_1, \theta_2) = (1 - \cos(\theta_1 - \theta_2)). \tag{2.13}$$

2.2 Most Common Circular Distributions

There are various distributions for circular data, such as uniform distribution, Cardioid distribution, wrapped normal distribution (WN), wrapped Cauchy distribution (WC) and von Mises distribution (vM). These distributions can be divided as unimodal-multi modal, symmetric-asymmetric, univariate-multivariate. Moreover, the distributions developed for circular data may be discrete or continuous. A circular probability density function (pdf) should satisfy three basic conditions;

1. $f(\theta) \geq 0$;
2. $\int_0^{2\pi} f(\theta)d\theta = 1$;
3. $f(\theta) = f(\theta + k2\pi)$ for any integer k.

Last one is the periodicity. In this section, three of these distributions will be given. These are the most common distribution in circular methodology.

2.2.1 The Von Mises (Circular Normal) Distribution

Probability density function for vM distribution is defined as

$$f(\theta|\mu, \kappa) = [2\pi I_0(\kappa)]^{-1} \exp\{\kappa \cos(\theta - \mu)\}, \quad 0 \leq \theta < 2\pi, \quad \kappa \geq 0 \qquad (2.14)$$

where μ is the mean direction and the concentration parameter is κ. Here, $I_0(\kappa)$ is the modified Bessel function of order zero and the first kind (Abramowitz and Stegun, 1965) and it is equal to

$$I_0(\kappa) = (2\pi)^{-1} \int_0^{2\pi} \exp(\kappa \cos(\phi))d\phi. \qquad (2.15)$$

$I_0(\kappa)$ also can be a calculated by using following expression;

$$I_0(\kappa) = \sum_{r=0}^{\infty} (r!)^{-2} (\frac{1}{2}\kappa)^{2r}. \qquad (2.16)$$

It was introduced by von Mises (1918). This distribution is symmetric about the direction μ as well as $\mu + \pi$ by the properties of the cosine function. The density has the maximum value at $\theta = \mu$ because the maximum value of cosine function is equal 1 and to satisfy this value the cosine function should be evaluated at 0, in other words μ is the modal direction. The minimum value of the cosine function is -1 and when the cosine function is evaluated at $\mu \pm \pi$, the density has the minimum value. Therefore, $\mu \pm \pi$ is the anti-modal direction of vM distribution. This distribution is a unimodal distribution. When concentration parameter κ is equal to 0, it reduces to uniform distribution and for small κ, the distribution can be approximated by cardioid distribution which is another circular distribution. Moreover, vM distribution can be approximated to wrapped normal (WN) and wrapped Cauchy (WC) distribution which will be discussed in next section. As κ increases, the distribution converges to a point at μ. Circular dispersion can be calculated for this distribution by using

$$\delta = [\kappa A_1(\kappa)]^{-1} \tag{2.17}$$

where $A_1(\kappa)$ is the mean resultant length and it is equal to $I_1(\kappa)/I_0(\kappa)$. Here $I_1(\kappa)$ denotes the modified Bessel function of order one and the first kind. In order to calculate values of the modified Bessel function of order p and the first kind, the following expression can be used ;

$$I_p(\kappa) = \sum_{r=0}^{\infty}[(r+p)!r!]^{-1}(\frac{1}{2}\kappa)^{2r+p},\ p = 1, 2, ... \tag{2.18}$$

2.2.2 The Wrapped Cauchy Distribution

Another symmetric and unimodal distribution for circular data is WC distribution. This distribution can be obtained by wrapping the Cauchy distribution on the real line around the circle and it was presented by Lévy (1939). Probability density function for WC distribution is defined as

$$f(\theta|\mu,\rho) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}, \quad 0 \le \theta < 2\pi, \quad 0 \le \rho < 1 \tag{2.19}$$

where μ is the mean direction and ρ is mean resultant length. As ρ increases, the distribution becomes concentrated at a point μ and as ρ approaches to 0, it reduces to uniform distribution. When modeling unimodal and symmetric data is concerned, this distribution can be used as an alternative to the vM distribution for suitable the dispersion parameters.

Circular dispersion can be obtained for this distribution as

$$\delta = (1-\rho^2)/(2\rho^2). \tag{2.20}$$

2.2.3 The Projected Normal Distribution

Let a random vector $\mathbf{X} = (X_1, X_2)'$ has a bivariate normal distribution, with mean μ and covariance matrix Σ. Let $\mathbf{U} = \frac{\mathbf{X}}{\|\mathbf{X}\|}$. Then the random unit vector $\mathbf{U}$ is a point on the unit circle and has a projected normal(PN) distribution (Small, 1996; Mardia et. al., 2000) with the same parameters which is denoted as $PN_2(\mu, \Sigma)$. Probability density function for two dimensional PN distribution is defined as

$$f(\theta|\mu,\Sigma) = \frac{\phi_2(\mu_1,\mu_2;0,\Sigma)}{C(\theta)} + \frac{aD(\theta)\Phi_1(D(\theta))\phi_1(aC(\theta)^{-\frac{1}{2}}(\mu_1\sin\theta-\mu_2\cos\theta))}{C(\theta)} \tag{2.21}$$

$a = (\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}$

$C(\theta) = a^2(\sigma_2^2\cos^2\theta - \rho\sigma_1\sigma_2\sin 2\theta + \sigma_1^2\sin^2\theta)$

$D(\theta) = a^2C(\theta)^{-\frac{1}{2}}(\mu_1\sigma_2(\sigma_2\cos\theta - \rho\sigma_1\sin\theta) + \mu_2\sigma_1(\sigma_1\sin\theta - \rho\sigma_2\cos\theta)),$

where ϕ_2 and Φ_2 denote the 2-dimensional probability density function and cumulative density function of standard normal distributions, respectively. $\mu = (\mu_1, \mu_2)'$ is mean vector and Σ is covariance matrix as given below

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where σ_1^2 and σ_2^2 denote the variances of X_1 and X_2 and ρ is the correlation coefficient between the two variables.

2.3 Regression Models

In this section, the theory of regression models for the circular variable will be described. In the literature, there are several regression models when the modeling with a circular variable is of concern. These regression models can be divided into three different groups. The first one is circular-circular regression model in which both response and explanatory variables are circular. The second one is circular-linear regression model that can be used to investigate the relationship between linear response variable and circular explanatory variables. The last one is linear-circular regression model, in which the variation of circular response variable can be explained in terms of some linear covariates. These regression models are generally developed under the von Mises distribution assumptions due to its established theoretical framework and some key properties for statistical inference. Our interest lies in linear-circular regressions. The variation of a circular response variable can be explained in a number of ways in terms of some linear covariates. For example, the mean direction or the dispersion of circular response variable can be modeled in terms of some explanatory variables, or we can model both the mean direction and the dispersion of circular response variable. In this study, the variation of circular response variable will be explained by modeling mean direction in linear-circular regression model. Other cases will not be discussed further.

2.3.1 Linear-Circular Regression Models

As mentioned above, the model in which the response variable is circular and explanatory variables are linear is called linear-circular regression model. In this model, the circular response variable is regressed on to one or more linear covariates and the variation of circular response can be investigated by modeling either its mean direction, concentration or both. Our interest lies in mean modeling.

2.3.1.1 Modeling the Mean Direction

In this section, regression model in which the mean direction is modeled in terms of some covariates will be given. Let $\theta_1, ..., \theta_n$ be a random sample of circular observations from von Mises distribution with mean directions μ_i's and concentration parameters κ. That is $\theta_i \sim vM(\mu_i, \kappa)$ for $i = 1, ..., n$. $\boldsymbol{X}_i$ be the vector of linear explanatory variables for the ith case. Firstly, Gould (1969) developed a regression model in which the structure

$$\mu_i = \mu_0 + \sum \boldsymbol{\beta}_i \boldsymbol{X}_i, \; i = 1, ..., n, \tag{2.22}$$

was considered for the mean direction, where, μ_0 is any constant, $\boldsymbol{X}_i$ is the vector of explanatory variables for the ith case and $\boldsymbol{\beta}_i$ is the vector of regression coefficients. The maximum likelihood estimates of the model parameters were obtained by using some iterative methods. The maximum likelihood estimations are equal to least squares estimations for this model. The likelihood function of this model has infinitely many equally high peaks, which lead to some problems in defining maximum likelihood estimators. This model was discussed by Laycock (1975) and some optimal designs were described for the related model. As an alternative to Gould's model, a different approach was proposed by Johnson and Wehrly (1978) for the case of only one linear covariate by using the joint distribution of the circular response variable and linear explanatory variable. They proposed the structure

$$\mu_i = \mu_0 + 2\pi F(x_i), \; i = 1, ..., n, \tag{2.23}$$

for the mean direction of von Mises distribution, where $F(x)$ is an exactly determined cumulative distribution function. In this model, the mean direction μ and concentration parameter κ can be estimated directly. This alternative approach was extended by Fisher and Lee (1992) to the case in which there are multiple linear covariates by using a general link function. For the mean direction, the structure

$$\mu_i = \mu_0 + g(\boldsymbol{\beta}^T \boldsymbol{X}_i), \; i = 1, ..., n, \tag{2.24}$$

was suggested by them, where $\boldsymbol{\beta}$ is vector of regression coefficients and $\boldsymbol{X}_i$ is the vector of predictors for the ith case and g is a general link function. They mapped the real line onto the unit circle by using the link function and just monotone link function was considered. The function g should be monotone and range from $-\pi$ to π with $g(0) = 0$. Some useful link functions are $g(x) = 2\pi F(x)$ where F(x) is a cdf and $g(x) = 2\tan^{-1}(x)$. Maximum likelihood estimates of model parameters can be obtained by using some iterative methods, however, when there are more than a few explanatory variables, it is difficult to work with the likelihood. George and Ghosh (2006) suggested a semi-parametric Bayesian approach to linear-circular regression. In this approach, nonparametric distribution such as Dirichlet process can be used as prior distribution for the regression coefficients. The projected normal distribution is another distribution which is used to propose an alternative approach for modeling the mean direction when the response variable is circular and the covariates are linear. In this alternative approach, a linear function of explanatory variables are used to explain the variation of the mean vector of the bivariate normal distribution. Some procedures can be used for obtaining the maximum likelihood estimations of model parameters.

CHAPTER 3

LONGITUDINAL CIRCULAR DATA

In this chapter, some basic properties of longitudinal data, an overview of the considered longitudinal circular random intercept model and Bayesian analysis of the model are presented.

3.1 Basis of Longitudinal Data

Longitudinal data occur when repeated observations are taken from the same individual through time. The longitudinal structure enables us to explore the change for each subject in response variable over time. Additionally, differences among subjects can be investigated in a repeated measurements study. In a longitudinal study, the primary objective is to assess the change within an individual over time and determine the covariates that affect the change (Fitzmaurice et al., 2004). There are three different sources of variation in longitudinal data. The first one is the variation of response variable among subjects, the second one is the variation of response variable over time within subjects and the next one is measurement error. Longitudinal data structures are mainly divided into three groups in Ilk (2008) as unconstrained, constrained and fully constrained. In unconstrained data, the number of time points for subjects is not equal to each other and repeated measurements are not taken from subjects on common time points. In constrained data, the number of time points is the same for all subjects but time points are not common. In fully constrained data, the number of time points is the same for all subjects and measurements are taken from each subject on the same time points. Longitudinal data can be constructed either prospectively or retrospectively. In the first way, repeated measurements from the same subject are taken in time, on the other hand, in the latter historical records are used to collect longitudinal data. Variables in longitudinal data can be divided as time dependent covariates, time independent covariates and response variable. Although, time dependent covariates can take different values in time, time independent covariates stay the same over time. These covariates can be continuous, discrete or categorical. Similarly, response variable in longitudinal data can be discrete, continuous or categorical as well as univariate or multivariate. Models built for continuous response may not be suitable for discrete or categorical response, thereby the choice of analysis is an important issue in longitudinal data and it depends on the variables in data set and design of data. Whether response variable is univariate or multivariate also plays important role in determining the models. Longitudinal data sets can be constructed either in long form or wide form and it can be reshaped from the long form to the wide form or vice versa.

A brief illustration of univariate longitudinal circular structure is presented before proceeding to longitudinal circular models. Assume that circular measurements are taken repeatedly from N subjects over time. Then, θ_{ij} indicates the circular response variable for the i^{th} subject on j^{th} time point. For each subject, grouping the circular response variable into an n_i x 1 response vector is appropriate as following;

$$\boldsymbol{\theta}_i = \begin{pmatrix} \theta_{i1} \\ \theta_{i2} \\ \vdots \\ \theta_{in_i} \end{pmatrix}, i = 1, ..., N \tag{3.1}$$

Since the number of time points for each subject may not be equal to each other, it is denoted by n_i for subject i. When all subjects have the same number of repeated measures, there is no need to include the index *i* in n_i. Moreover, associated $p \times 1$ linear covariates vector is denoted by

$$\boldsymbol{X}_{ij} = \begin{pmatrix} X_{ij1} \\ X_{ij2} \\ \vdots \\ X_{ijp} \end{pmatrix}, i = 1, ..., N;\ j = 1, ..., n_i. \tag{3.2}$$

Note that X_{ijp} denotes the value of p^{th} covariate for subject i on j^{th} time point. These *p* covariates may change over time or not change as mentioned earlier.

3.2 A Longitudinal Circular Random Effects Model

Let $\theta_{ij} \in [-\pi, \pi]$ be a circular random variable in a repeated measurements study and let the number of time points for each subject be the same and common. That is, $i = 1, ..., n,\ j = 1, ..., m$ and θ_{ij} denotes the measure of a circular response for subject *i* on time point *j*. Then, we consider the following random intercept model (which we will denote by LCRIM)

$$\begin{aligned} &\theta_{ij} = 2\arctan(b_{0i} + \boldsymbol{\beta}\boldsymbol{X}_{ij}) + \epsilon_{ij},\ i = 1, ..., n,\ j = 1, ..., m \\ &\epsilon_{ij} \overset{ind}{\sim} vM(0, \kappa) \end{aligned} \tag{3.3}$$

where b_{0i} is subject specific random intercept for subject *i* which has a normal distribution with mean μ_{b_0} and variance $\sigma^2_{b_0}$. $\boldsymbol{X}_{ij}$ are linear covariates, $\boldsymbol{\beta}$ are regression coefficients, and κ is the concentration parameter. b_{0i} and ϵ_{ij}'s are assumed independent. It is also assumed that θ_{ij} and $\theta_{ij}{}'$ are conditionally independent given a subject-specific random intercept.

An equivalent hierarchical representation is given as follows:

$$\begin{aligned} \theta_{ij}|b_{0i} &\sim vM(\mu_{ij}, \kappa),\ i = 1, ..., n,\ j = 1, ..., m \\ \mu_{ij} &= g(b_{0i} + \boldsymbol{\beta}\boldsymbol{X}_{ij}) \\ b_{0i} &\sim N(\mu_{b_0}, \sigma^2_{b_0}). \end{aligned} \tag{3.4}$$

with monotone increasing link function such as $g(x) = 2\arctan(x)$.

Interpretation of the regression coefficients in a circular model is different than those in a regression model defined on Euclidean space and thus requires special attention. First of all, for the correct interpretation, the initial direction and the direction of rotation should be determined clearly. To illustrate, let the initial direction be North and the direction of rotation be clockwise. Then, a positive β_p means that an increase in the p^{th} covariate will lead to a clockwise advance from North in the directional response. In other words, when there is one unit change in the p^{th} covariate, there will be a change in $\tan(\frac{\theta_{ij}}{2})$ as β_p.

3.3 Bayesian Analysis of LCRIM

In this section, a Bayesian analysis of longitudinal circular random intercept model is presented. In this framework, first, prior distributions are specified for all unknown parameters of the model. Second, all these distributions are combined into the joint posterior distribution. Afterwards the full conditional distribution are derived from the joint posterior distribution for each parameter. Finally, Markov Chain Monte Carlo (MCMC) methods are employed to obtain the random samples from the marginal posterior distribution for each parameter of model and the parameter estimates. The directed graphical model (DAG) for LCRIM is presented in Appeddix A.

The longitudinal circular random intercept model given in Section 3.2 has the parameter space consisting of $(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau)$ where κ is the concentration parameter, $\boldsymbol{\beta} = \{\beta_p; p = 1, ..., k\}$ is a vector of regression coefficients, $\tau = \frac{1}{\sigma^2_{b_0}}$ and μ_{b_0} is the mean of the assumed distribution of the random intercept. We choose to use the standard prior distributions; Normal distributions for the parameters in $\mathbb{R}$ and Gamma distributions for the parameters in $\mathbb{R}^+$. The prior distributions are $\beta_p \sim N(\mu_{\beta_p}, \sigma^2_{\beta_p})$, $\kappa \sim Ga(a_\kappa, b_\kappa)$, $\mu_{b_0} \sim N(\mu_0, \sigma^2_0)$, and $\tau \sim Ga(a_\tau, b_\tau)$ with known hyperpriors.

Let $D_{obs} = \{\boldsymbol{\theta}, \boldsymbol{X}\}$ and $D_{comp} = \{\boldsymbol{\theta}, \boldsymbol{X}, b_0\}$ be the observed and complete data matrices respectively. Complete data matrix includes both the observables and unobservables (latent variable random intercept). The joint posterior distribution of all unknown quantities given the observed data D_{obs} is presented as

$$\begin{aligned} f(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau, b_0|D_{obs}) =& \frac{f(D_{obs}, b_0|\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau) f(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau)}{f(D_{obs})} \\ =& \frac{f((D_{comp}|\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau) f(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau)}{f(D_{obs})} \\ \propto& f(D_{comp}|\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau) f(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau) \end{aligned} \tag{3.5}$$

where $f(D_{comp}|\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau)$ and $f(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau)$ denote the complete data likelihood function and the joint prior distribution of the parameters respectively. Since the denominator does not include κ, $\boldsymbol{\beta}$, μ_{b_0} or τ, it is not really important and the proportion can be used instead of equality. Then, the joint posterior density of variables can be obtained by multiplying the complete data likelihood by the prior distributions for each variable. The complete data likelihood function for LCRIM can be written as below:

$$L(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau|D_{comp}) = \prod_{i=1}^{n} \left(\prod_{j=1}^{m} f(\theta_{ij}|b_{0i}, \boldsymbol{\beta}, \kappa) \right) f(b_{0i}|\mu_{b_0}, \tau), \tag{3.6}$$

where $f(\theta_{ij}|b_{0i}, \boldsymbol{\beta}, \kappa)$ denotes the conditional circular pdf (vM) and $f(b_{0i}|\mu_{b_0}, \tau)$ is the prior distribution for b_{0i} (Normal distribution). Then,

$$\begin{aligned} L(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau|D_{comp}) =& \prod_{i=1}^{n} \left(\prod_{j=1}^{m} [2\pi I_0(\kappa)]^{-1} \exp\{\kappa \cos(\theta_{ij} - 2\arctan(b_{0i} + \boldsymbol{\beta}\boldsymbol{X}_{ij}))\} \right) \\ & \frac{\sqrt{\tau}}{\sqrt{2\pi}} exp\{-\frac{\tau(b_{0i} - \mu_{b_0})^2}{2}\} \end{aligned} \tag{3.7}$$

After a little bit of algebra,

$$L(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau|D_{comp}) =[2\pi I_0(\kappa)]^{-mn} exp\{\kappa \sum_{i=1}^{n}\sum_{j=1}^{m} \cos(\theta_{ij} - 2\arctan(b_{0i} + \boldsymbol{\beta X}_{ij}))\}$$

$$(\frac{\tau}{2\pi})^{\frac{n}{2}} exp\{-\tau \sum_{i=1}^{n} \frac{(b_{0i} - \mu_{b_0})^2}{2}\}. \tag{3.8}$$

Then, the joint posterior density of all unknown variates is presented as

$$f(\kappa, \boldsymbol{\beta}, \mu_{b_0}, \tau, b_0|D_{obs}) \propto [2\pi I_0(\kappa)]^{-mn} exp\{\kappa \sum_{i=1}^{n}\sum_{j=1}^{m} \cos(\theta_{ij} - 2\arctan(b_{0i} + \boldsymbol{\beta X}_{ij}))\}$$

$$(\frac{\tau}{2\pi})^{\frac{n}{2}} exp\{-\tau \sum_{i=1}^{n} \frac{(b_{0i} - \mu_{b_0})^2}{2}\} \frac{1}{\sqrt{2\pi}\sigma_\beta} exp\{-\frac{(\boldsymbol{\beta} - \mu_\beta)^2}{2\sigma_\beta^2}\}$$

$$\frac{b_\kappa^{a_\kappa}}{\Gamma(a_\kappa)} \kappa^{a_\kappa - 1} exp\{-\kappa b_\kappa\} \frac{1}{\sqrt{2\pi}\sigma_0} exp\{-\frac{(\mu_{b_0} - \mu_0)^2}{2\sigma_0^2}\}$$

$$\frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)} \tau^{a_\tau - 1} exp\{-\tau b_\tau\} \tag{3.9}$$

This joint posterior density should be integrated out for all the other parameters to obtain the marginal posterior density for each parameter. However, the joint posterior density is very complicated and not available in closed form, thereby, MCMC methods (e.g. Gibbs sampling) are used to draw a random sample from the marginal posterior density. Gibbs sampling algorithm uses the full conditional distributions of the parameters to draw samples from the marginal posterior densities of the parameters.

The full conditional density for each variable can be derived from the joint posterior density given in equation 3.9.

Full conditional distribution for μ_{b_0},

$$f(\mu_{b_0}|\boldsymbol{\beta},\kappa,\tau,b_0,D_{obs}) = \frac{f(\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,b_0|D_{obs})}{f(\boldsymbol{\beta},\kappa,\tau,b_0|D_{obs})} \\ \propto f(\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,b_0|D_{obs}). \tag{3.10}$$

The denominator does not contain μ_{b_0} and is a fixed value for a given set of ($\boldsymbol{\beta}$, κ, τ, b_0) at a given Gibbs iteration and thus the proportion can be used instead of equality. Then, the full conditional distribution for μ_{b_0} is represented as

$$f(\mu_{b_0}|\boldsymbol{\beta},\kappa,\tau,b_0,D_{obs}) \propto (\frac{\tau}{2\pi})^{\frac{n}{2}} exp\{-\frac{\tau}{2}\sum_{i=1}^{n}(b_{0i}-\mu_{b_0})^2\} \\ \frac{1}{\sqrt{2\pi}\sigma_0} exp\{-\frac{(\mu_{b_0}-\mu_0)^2}{2\sigma_0^2}\} \\ \propto exp\{-\frac{\tau}{2}\sum_{i=1}^{n}(b_{0i}-\mu_{b_0})^2 - \frac{(\mu_{b_0}-\mu_0)^2}{2\sigma_0^2}\} \tag{3.11}$$

After a little bit of algebra

$$f(\mu_{b_0}|\boldsymbol{\beta},\kappa,\tau,b_0,D_{obs}) \propto exp\{-\frac{1}{2}(1+n\tau)(\mu_{b_0}-\frac{\tau\sum b_{0i}}{1+n\tau})^2\} \tag{3.12}$$

Hence, the full conditional distribution for μ_{b_0} is a normal distribution with mean $\frac{\tau\sum b_{0i}}{1+n\tau}$ and variance $(1+n\tau)^{-1}$. μ_0 and σ_0^2 are considered as 0 and 1 respectively for simplicity.

Full conditional distribution for τ,

$$f(\tau|\boldsymbol{\beta},\kappa,\mu_{b_0},b_0,D_{obs}) = \frac{f(\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,b_0|D_{obs})}{f(\boldsymbol{\beta},\kappa,\mu_{b_0},b_0|D_{obs})} \\ \propto f(\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,b_0|D_{obs}). \tag{3.13}$$

The denominator again does not include τ and the proportion can be used instead of equality. Then, the full conditional distribution for τ is represented as

$$
\begin{aligned}
f(\tau|\boldsymbol{\beta}, \kappa, \mu_{b_0}, b_0, D_{obs}) \propto & \tau^{\frac{n}{2}} exp\{-\tau \sum_{i=1}^{n} \frac{(b_{0i}-\mu_{b_0})^2}{2}\} \frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)} \tau^{a_\tau - 1} exp\{-\tau b_\tau\} \\
& \propto \tau^{\frac{n}{2}+a_\tau-1} exp\{-\tau (b_\tau + \frac{1}{2}\sum_{i=1}^{n}(b_{0i}-\mu_{b_0})^2\}
\end{aligned} \tag{3.14}
$$

which can be recognized as the kernel of a gamma distribution with shape parameter $\frac{n}{2}+a_\tau$ and scale parameter $b_\tau + \frac{1}{2}\sum_{i=1}^{n}(b_{0i}-\mu_{b_0})^2$.

Joint full conditional distribution for β and κ,

$$
\begin{aligned}
f(\boldsymbol{\beta}, \kappa|\mu_{b_0}, \tau, b_0, D_{obs}) &= \frac{f(\boldsymbol{\beta}, \kappa, \mu_{b_0}, \tau, b_0|D_{obs})}{f(\mu_{b_0}, \tau, b_0|D_{obs})} \\
& \propto f(\boldsymbol{\beta}, \kappa, \mu_{b_0}, \tau, b_0|D_{obs}).
\end{aligned} \tag{3.15}
$$

The denominator again does not include $\boldsymbol{\beta}$ and κ so the proportion can be used instead of equality. Then, the joint full conditional distribution for $\boldsymbol{\beta}$ and κ is represented as

$$
\begin{aligned}
f(\boldsymbol{\beta}, \kappa|\mu_{b_0}, \tau, b_0, D_{obs}) \propto & [2\pi I_0(\kappa)]^{-mn} exp\{\kappa \sum_{i=1}^{n}\sum_{j=1}^{m} \cos(\theta_{ij} - 2\arctan(b_{0i} + \boldsymbol{\beta}\boldsymbol{X}_{ij}\} \\
& \frac{1}{\sqrt{2\pi}\sigma_\beta} exp\{-\frac{(\boldsymbol{\beta}-\mu_\beta)^2}{2\sigma_\beta^2}\} \frac{b_\kappa^{a_\kappa}}{\Gamma(a_\kappa)} \kappa^{a_\kappa - 1} exp\{-\kappa b_\kappa\}
\end{aligned} \tag{3.16}
$$

which is very complicated to recognize as any closed form distribution. MCMC methods (e.g. Gibbs sampling algorithm, Metropolis-Hastings (M-H) algorithms) can be used to deal with this complicated situation. For computationally more efficient block sampling algorithm, joint full conditional distribution was employed for these parameters. In Bayesian approach, latent random intercept can be viewed as an unknown parameter and the following full conditional distribution is employed to sample for random intercept,

$$\begin{aligned} f(b_{0i}|\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,D_{obs}) &= \frac{f(\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,b_0|D_{obs})}{f(\boldsymbol{\beta},\kappa,\mu_{b_0},\tau|D_{obs})} \\ &\propto f(\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,b_0|D_{obs}),\ for\ i=1,..,n. \end{aligned} \tag{3.17}$$

Since the denominator again does not involve b_{0i}, this full conditional distribution can be expressed as

$$\begin{aligned} f(b_{0i}|\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,D_{obs}) &\propto exp\{\kappa\sum_{i=1}^{n}\sum_{j=1}^{m}\cos(\theta_{ij}-2\arctan(b_{0i}+\boldsymbol{\beta}\boldsymbol{X}_{ij}\} \\ &exp\{-\frac{\tau}{2}\sum_{i=1}^{n}(b_{0i}-\mu_{b_0})^2\},\ for\ i=1,..,n. \end{aligned} \tag{3.18}$$

In order to obtain the predicted data, a posterior predictive distribution is used as follows. Let $\theta^{pred}=(\theta_1^{pred},...,\theta_n^{pred})$ and $\theta^{obs}=(\theta_1^{obs},...,\theta_n^{obs})$ be predicted (unknown) and observed data, respectively. Predicted data are obtained from posterior predictive distribution as follows

$$f(\theta_i^{pred}|\theta^{obs}) = \int f(\theta_i^{pred}|\Omega)f(\Omega|\theta^{obs})d\Omega,\ for\ i=1,...,n, \tag{3.19}$$

where $f(\theta^{pred}|\Omega)$ is the posterior predictive density of the data which is the density evaluated at θ^{pred} given the parameter vector $\Omega=(\boldsymbol{\beta},\kappa,\mu_{b_0},\tau,b_0)$ and $f(\Omega|\theta^{obs})$ is the joint posterior density of the all parameters given the observed data θ^{obs}.

For each parameter, sampling algorithms within the Gibbs sampling are presented in Table 3.1.

Table 3.1: The Sampling Algorithms

Latent/ Parameter	Sampling Method
β	Adaptive Metropolis Block
κ	Adaptive Metropolis Block
b_{0i}	Adaptive Metropolis
μ_{b_0}	Direct Sampling
τ	Direct Sampling

CHAPTER 4

MODEL ASSESSMENT COMPARISON AND SELECTION

With the advances in technology, more circular data emerge in several different areas requiring complex circular data modeling and accordingly a lot of statistical models that can be used to analyze circular data have been developed. However, research on model comparison and selection in complex circular modeling literature is rather in its infancy: e.g. studies on their performances are inadequate. In this thesis, we focus on predictive model selection. Below we list the methods in circular literature that can be used for predictive model assessment, comparison and selection.

- GAIC: Developed by Ye (1998); adapted for circular data and its performance is evaluated for distinguishing between standard circular-circular and circular-piecewise regressions by Rueda et al. (2016).

- DIC: Developed by Spiegelhalter et al. (2002); its variants for random effects models were developed by Celeux et al. (2006); original DIC is adapted for circular models and its performance is evaluated by Ravindran and Ghosh (2011) for selecting the best wrapped model.

- CPO: Conditional predictive ordinate developed by Geisser and Eddy (1979); asymptotically equivalent to AIC; used by Nunez-Antonio et al. (2011) for missing and by McMillan et al. (2013) for repeated measures circular data modeling.

- AIC: Penalty (number of parameters) is not clear for complex models such as mixed effects models; however it is still in use for comparing models for longitudinal/clustered circular data.

- Average circular residual: Defined as the average cosine distances between the observed and predicted circular data; used by Maruotti et al. (2016) to compare the prediction abilities of complex longitudinal models (PN and VM models with random effects with HMM vs. Gaussian).

- CPD (Circular Predictive Discrepancy): Developed by Ravindran and Ghosh (2011); it is based on the decision theoretic model choice approach of Gelfand and Ghosh (1998) for linear data; its performance is evaluated in selecting the best wrapped distribution.

In this thesis, we focus on the prediction error based methods (namely CPD), propose new methods and compare the proposed method with CPD. In this chapter CPD measure and proposed model assessment and model comparison criteria are presented in detail.

4.1 Circular Predictive Discrepancy Measures

4.1.1 Circular Predictive Discrepancy-Type 1

Loss functions for circular data are quite different from those of linear data. When it is desired to define a loss function for circular data, the nature of circular data must also be taken into account. In the context of circular data, although the theory of loss functions is a rather neglected issue of statistics, Ravindran, et. al. (2011) defined a loss function measuring the difference between the predicted and the observed circular measurements. They used the theory of decision (Gelfand and Ghosh, 1998) to define this loss function and called it Absolute Predicted Errors (we will call it APE_1). It is given by

$$APE_1 = \sum_{i=1}^{n} min(|\theta_i^{pred} - \theta_i^{obs}|, 2\pi - |\theta_i^{pred} - \theta_i^{obs}|). \tag{4.1}$$

They proposed a circular predictive discrepancy-type 1 as a model selection criteria as follows

$$CPD_1 = E[APE_1|\theta^{obs}]. \quad (4.2)$$

We will call this Circular Predictive Discrepancy-Type 1 and denote it by CPD_1. This criterion is based on minimizing a predictive loss. They used posterior predictive distribution to calculate CPD_1. Although loss functions for linear data can be broken into two terms as a penalty term and a goodness of fit term, CPD_1 can not be split into these two terms. Lower CPD_1 indicates a better fit. Note that, APE_1 for longitudinal circular data would be as follows

$$APE_1 = \sum_{i=1}^{n} \sum_{j=1}^{m} min(|\theta_{ij}^{pred} - \theta_{ij}^{obs}|, 2\pi - |\theta_{ij}^{pred} - \theta_{ij}^{obs}|). \quad (4.3)$$

4.1.2 Circular Predictive Discrepancy-Type 2

Jammalamadaka and SenGupta (2001) presented a definition of circular distance between any two points on the circle as follows. Letting α and β be two angles

$$d(\alpha, \beta) = 1 - cos(\alpha - \beta). \quad (4.4)$$

It is clear that this circular distance is a monotone increasing function of angle between two points K and L points in Figure 4.1: e.g. θ. Since cosine function can take values between -1 and 1, the minimum and maximum values that circular distance between any two points can take are equal to 0 and 2, respectively.

We employ the circular distance suggested by Jammalamadaka and SenGupta (2001) to introduce a model comparison criterion for circular data and call it Circular Predictive Discrepancy-Type 2 (CPD_2). For longitudinal circular data, we define it as follows

$$CPD_2 = E[APE_2|\theta^{obs}], \quad (4.5)$$

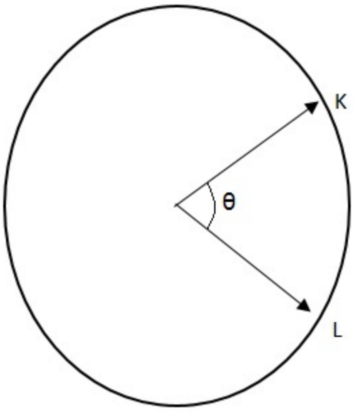

Figure 4.1: Circular distance between K and L points. Adapted from Jammalamadaka and SenGupta (2001).

where

$$APE_2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (1 - \cos(\theta_{ij}^{pred} - \theta_{ij}^{obs})). \tag{4.6}$$

This is the total angular distance between the observed and the predicted angles (points on the unit circle). Any model with lowest CPD_2 is selected as the best fitting model.

4.2 Plug-in Absolute Predicted Errors

4.2.1 Plug-in Absolute Predicted Errors-Type 1

We use the theoretic framework of loss function of Ravindran et. al. (2011) to define a model comparison criterion directly for longitudinal circular data and call it Plug-in Absolute Predicted Errors-Type 1 ($PAPE_1$). It can be defined as follows

$$PAPE_1 = \sum_{i=1}^{n} \sum_{j=1}^{m} min(|E[\theta_{ij}^{pred}|\theta_{ij}^{obs}] - \theta_{ij}^{obs}|, 2\pi - |E[\theta_{ij}^{pred}|\theta_{ij}^{obs}] - \theta_{ij}^{obs}|), \tag{4.7}$$

where $E[\theta_{ij}^{pred}|\theta_{ij}^{obs}]$ is posterior predictive mean for subject (cluster) *i* on time point *j* circular outcome. Again, posterior predictive distribution is employed to obtain posterior predictive mean and this model comparison criterion is depending on minimizing a predictive loss. A model with lowest $PAPE_1$ is selected as a best fitting model.

4.2.2 Plug-in Absolute Predicted Errors-Type 2

We employ posterior predictive mean in APE_2 to obtain a model comparison criterion directly for longitudinal circular data and call it Plug-in Absolute Predicted Errors-Type 2 ($PAPE_2$). It is given by

$$PAPE_2 = \sum_{i=1}^{n}\sum_{j=1}^{m}(1 - \cos(E[\theta_{ij}^{pred}|\theta_{ij}^{obs}] - \theta_{ij}^{obs}) \tag{4.8}$$

where $E[\theta_{ij}^{pred}|\theta_{ij}^{obs}]$ again denotes the posterior predictive mean for subject i's j^{th} circular predicted value. In order to obtain posterior predictive mean, posterior predictive distribution should be used. The logic for this model comparison criterion is the same as CPD_2 mentioned earlier. A model is preferred which gives the lowest value of $PAPE_2$.

CHAPTER 5

SIMULATION STUDY

In this chapter, an extensive Monte Carlo (MC) simulation study is presented under various realistic balanced longitudinal settings. Aims of this MC simulation study are to show the performance of the parameter estimation technique for the longitudinal circular random intercept model presented in Chapter 3 and to evaluate the performances of model selection and comparison criteria detailed in Chapter 4. A number of different simulation scenarios are controlled for the sample size, effect size and intraclass correlation coefficient value, ICC=0.20, 0.50, 0.88. Choice of ICC values represent low, mild and high intraclass correlation, respectively. R language and environment (R Development Core Team, 2017) is used to simulate data sets for the different scenarios. In order to implement MCMC scheme, OpenBUGS which is an open source software for Bayesian statistics is employed. R programming language and OpenBUGS program are integrated to carry out all analyses in this thesis. The number of repeated measurements which is the same for all subjects is assumed to be equal to five for all simulation scenarios and the time points that measurements are taken are common (i.e. equally spaced) for each subject. All Monte Carlo scenarios are repeated 100 times.

In order to investigate the performance of parameter estimations, Relative Bias (RB) and Mean Square Error (MSE) are used. Calculation of these measures are as follows

$$RB = \frac{E(\hat{\beta}) - \beta}{\beta}, \; MSE = E[(\hat{\beta} - \beta)^2], \tag{5.1}$$

where, β is true value for estimation of interest and $\hat{\beta}$ is the value for estimator of interest. If RB and MSE of parameter estimators do not change under model misspecification, then parameter estimators are robust to model misspecification. Frequency of selecting the true model and average stability coefficient (ASC) which will be detailed in Section 5.3.2 are used to evaluate the performances of model selection criteria. The posterior mean, standard deviation, 2.5th percentile, median, 97.5th percentile for parameters are also computed for each scenario.

Convergence diagnosis is an significant issue in MCMC studies and requires special attention. We used trace plots and Brooks-Gelman-Rubin (BGR) statistic (Brooks and Gelman, 1998) for convergence diagnostics and for determining burn-in (warm-up) period as well as size of MCMC samples to be used for final posterior inference. To calculate the BGR statistic, we used two different chains with different initial values. BGR statistic with the value less than 1.1 indicates convergence. This condition is satisfied for all scenarios. Another important issue in MCMC studies is to determine the number of samples that is needed for posterior inference. It is known that the more samples are saved after convergence, the more accurate will be the posterior estimates. Given a chain size, accuracy of the posterior estimates is determined by Monte Carlo (MC) error. MC errors less than 5% of the sample standard deviation indicates sufficient accuracy. All simulation scenarios are run until this rule is met.

This extensive simulation study is basically divided into two groups as parameter estimations (Section 5.1) and model assessment, comparison and selection (Section 5.2). Moreover, these sections are divided into two groups as specification of mean models and different distributions.

Here we give a brief description of ICC in the circular context. ICC is the proportion of between subject variability to the total variability containing error variance (uncertainty) and between-subject variability. Let θ_{ij} be circular observation for subject (cluster) *i* on time point *j*, $i = 1, ...n$, $j = 1, ..., m$ and then let,

$$\theta_{ij} = 2\arctan(b_{0i} + \boldsymbol{\beta X}_{ij}) + \epsilon_{ij}, \tag{5.2}$$

where $\epsilon_{ij} \sim vM(0, \kappa)$, $b_{0i} \sim N(\mu_{b_0}, \sigma^2_{b_0})$. Let σ^2_ϵ denotes the common error variance, namely $Var(\epsilon_{ij})$ which is given by

$$Var(\epsilon_{ij}) = [\kappa A_1(\kappa)]^{-1} \tag{5.3}$$

The calculation of ICC for $\tan(\frac{\theta_{ij}}{2})$ can be denoted by

$$ICC = \frac{\sigma_{b_0}^2}{\sigma_{b_0}^2 + [\kappa A_1(\kappa)]^{-1}}. \tag{5.4}$$

In the simulations, $\sigma_{b_0}^2$ and κ are chosen so that the simulations are controlled for ICC. For instance, for ICC=0.20 and $\kappa = 2$, $\sigma_{b_0}^2$ is calculated as 0.18 using equation 5.4. ICC is one of the most practical reliability measure that can be employed to evaluate absolute agreement or level of consistency. There are many different formulas for calculating the ICC (Haggard, 1958; Feldt, 1965; Bartko, 1966; Snedecor and Cochran, 1967; Winer, 1971; Hayes, 1973).

5.1 Parameter Estimations

In this section, performances of parameter estimators defined in Section 3.2 are investigated under correctly specified model and under model misspecification. Model misspecification is introduced *i.* in the mean model of the distribution, and *ii.* in distribution function specification.

5.1.1 Specification of Mean Models

The performance of parameter estimation technique is evaluated for linear-quadratic mean models (Section 5.1.1.1) and main-interaction mean models (Section 5.1.1.2).

5.1.1.1 Linear and Quadratic Mean Models

In this study, we consider the two types of true models (TM). These are denoted by TM1 and TM2 as follows:

TM1: $\mu_{ij} = 2\arctan(b_{0i} + \beta_1 x_{ij} + \beta_2 x_{ij}^2)$, TM2: $\mu_{ij} = 2\arctan(b_{0i} + \beta_1 x_{ij})$.

As seen, TM1 is a quadratic model whereas TM2 is a linear model. We also considered two different values for β_2. β_2 is set at 1.5 or 0.3 representing effective and relatively ineffective quadratic term to control the simulations for quadratic term effect size. The value of the concentration parameter, κ of vM distribution should be at least 2 so that some statistical techniques can be applied and various key approximations such as Bessel function approximations are possible in circular statistics (Fisher, 1993). For this reason, the true value of the concentration parameter, κ is set equal to 2 and this value is used for all scenarios. True parameter values are given in parentheses in subsequent tables. Just single time-dependent linear explanatory variable x_{ij} is used and produced by random draws from a standard normal distribution.

Fitted models denoted by M1 and M2 are as follows

M1: $\mu_{ij} = 2\arctan(b_{0i} + \beta_1 x_{ij} + \beta_2 x_{ij}^2)$

M2: $\mu_{ij} = 2\arctan(b_{0i} + \beta_1 x_{ij})$.

Tables show RB and MSEs of Bayesian estimators of the parameters of interest under different TM, ICC, and β_2. MSE values and true values of parameters are presented in parenthesis throughout this thesis . For Tables 5.1-5.6, data are generated from TM1. In these tables, Model=M1 lines determine the behaviour of the estimators when fitted model correctly specifies the underlying true model whereas Model=M2 lines determine those under mean model misspecification. For Tables 5.7-5.9, TM2 is used to generate the data. Model=M2 lines indicate the behaviour of the estimators for fitting true model and Model=M1 lines denote the results when the fitted model is misspecified.

According to the results of Table 5.1 and Table 5.3, for low and mild ICC and high β_2, RB and MSE of all parameters converge to 0 fast when the mean model is correctly specified, as sample size increases. RB and MSE are at acceptable level irrespective of the sample size when fitted model is correctly specified. For these scenarios, parameter estimators are not robust to model misspecification

Table 5.1: RB and MSE values. True model = TM1, ICC = 0.20, β_2=1.5.

n	Model	$\beta_1(2.5)$	$\beta_2(1.5)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.18)$
20	M1	0.056(0.104)	0.062(0.093)	0.016(0.067)	-0.006(0.0178)	0.005(0.0001)
	M2	-0.093(0.161)	-	-0.299(0.396)	0.221(0.066)	0.002(0.0000)
50	M1	0.035(0.058)	0.035(0.054)	0.018(0.032)	0.017(0.009)	0.001(0.0002)
	M2	-0.124(0.161)	-	-0.306(0.392)	0.234(0.064)	-0.004(0.0001)
100	M1	0.012(0.015)	0.007(0.014)	-0.003(0.009)	0.0002(0.004)	-0.002(0.0002)
	M2	-0.140(0.144)	-	-0.312(0.396)	0.222(0.054)	-0.009(0.0002)
250	M1	0.004(0.008)	0.005(0.007)	-0.0001(0.006)	0.0009(0.001)	0.005(0.0002)
	M2	-0.147(0.142)	-	-0.323(0.421)	0.219(0.050)	-0.021(0.0005)
500	M1	0.001(0.005)	0.0006(0.004)	-0.0001(0.003)	-0.001(0.0007)	-0.0001(0.0003)
	M2	-0.154(0.154)	-	-0.328(0.429)	0.213(0.046)	-0.033(0.0012)

Table 5.2: RB and MSE values. True model = TM1, ICC = 0.20, β_2=0.3.

n	Model	$\beta_1(2.5)$	$\beta_2(0.3)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.18)$
20	M1	0.088(0.199)	-0.030(0.150)	-0.0005(0.075)	0.004(0.024)	0.004(0.0001)
	M2	0.041(0.128)	-	-0.007(0.074)	0.038(0.023)	0.004(0.0001)
50	M1	0.029(0.047)	-0.072(0.042)	0.003(0.021)	0.002(0.008)	0.005(0.0002)
	M2	0.007(0.040)	-	-0.004(0.021)	0.039(0.008)	0.004(0.0002)
100	M1	0.018(0.024)	-0.028(0.022)	-0.005(0.012)	0.008(0.004)	0.002(0.0002)
	M2	0.001(0.019)	-	-0.011(0.012)	0.048(0.006)	0.002(0.0002)
250	M1	0.002(0.007)	-0.028(0.007)	-0.004(0.006)	-0.002(0.002)	0.0006(0.0003)
	M2	-0.009(0.008)	-	-0.009(0.006)	0.039(0.003)	-0.0004(0.0003)
500	M1	0.006(0.004)	0.014(0.004)	0.001(0.002)	-0.001(0.0008)	0.0003(0.0003)
	M2	-0.005(0.0040)	-	-0.004(0.002)	0.041(0.002)	-0.0006(0.0003)

Based on the results of Table 5.2 and Table 5.4, in terms of RB and MSE, in which true β_2 is not that large, parameter inferences for random intercept variance are robust to model misspecification. This may be due to the small effect of quadratic term. That is, there is almost no difference between these two models in these scenarios. Parameter estimators of other parameters are not robust to model misspecification. As sample size increases, RB and MSE for all parameters decrease, when the model is both correctly specified and misspecified. Looking at the overall tables for low ICC

Table 5.3: RB and MSE values. True model = TM1, ICC = 0.50, β_2=1.5.

n	Model	$\beta_1(2.5)$	$\beta_2(1.5)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.72)$
20	M1	0.055(0.157)	0.076(0.151)	0.010(0.085)	-0.0007(0.054)	0.219(0.154)
	M2	-0.053(0.295)	-	-0.308(0.434)	0.383(0.231)	0.294(0.285)
50	M1	0.024(0.056)	0.031(0.054)	0.002(0.038)	0.009(0.024)	0.057(0.046)
	M2	-0.105(0.169)	-	-0.327(0.450)	0.362(0.163)	0.053(0.059)
100	M1	0.014(0.025)	0.023(0.025)	0.007(0.013)	0.005(0.011)	0.029(0.029)
	M2	-0.130(0.136)	-	-0.323(0.425)	0.362(0.145)	0.035(0.055)
250	M1	0.010(0.009)	0.009(0.009)	0.005(0.009)	0.0006(0.004)	0.014(0.009)
	M2	-0.137(0.131)	-	-0.316(0.406)	0.346(0.125)	-0.011(0.020)
500	M1	0.003(0.005)	0.003(0.005)	0.0004(0.003)	0.006(0.003)	0.009(0.006)
	M2	-0.145(0.140)	-	-0.322(0.417)	0.356(0.131)	-0.010(0.011)

Table 5.4: RB and MSE values. True model = TM1, ICC = 0.50, β_2=0.3.

n	Model	$\beta_1(2.5)$	$\beta_2(0.3)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.72)$
20	M1	0.118(0.224)	0.019(0.137)	0.037(0.119)	-0.007(0.077)	0.183(0.114)
	M2	0.069(0.147)	-	0.027(0.106)	0.061(0.067)	0.157(0.101)
50	M1	0.039(0.056)	-0.029(0.046)	0.006(0.033)	-0.013(0.028)	0.119(0.063)
	M2	0.011(0.044)	-	-0.002(0.032)	0.053(0.029)	0.113(0.063)
100	M1	0.016(0.025)	-0.078(0.023)	0.005(0.018)	0.007(0.014)	0.032(0.024)
	M2	-0.003(0.021)	-	-0.002(0.017)	0.071(0.017)	0.025(0.022)
250	M1	0.004(0.010)	-0.054(0.010)	0.006(0.007)	0.008(0.005)	0.025(0.011)
	M2	-0.012(0.011)	-	-0.0006(0.007)	0.074(0.009)	0.024(0.012)
500	M1	0.002(0.005)	-0.021(0.005)	-0.006(0.003)	0.005(0.003)	0.006(0.005)
	M2	-0.008(0.005)	-	-0.012(0.004)	0.073(0.008)	0.004(0.005)

(Table 5.2) and mild ICC (Table 5.4), RB and MSE quantities are at acceptable level irrespective of sample size and ICC values.

According to Table 5.5 when ICC is high, parameter estimators are not robust to model misspecification. Random effects variance has little bit high values for these quantities compared to low and mild ICC when the model is not correctly specified. As sample size increases, RB and MSE of the parameters decrease when correct model is fitted. For all parameters, when the model is correctly specified, RB and MSE measures are at acceptable level irrespective of the sample size.

Table 5.5: RB and MSE values. True model = TM1, ICC = 0.88, β_2=1.5.

n	Model	$\beta_1(2.5)$	$\beta_2(1.5)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(5.28)$
20	M1	0.009(0.015)	0.016(0.015)	-0.004(0.088)	0.009(0.019)	0.828(3.852)
	M2	0.002(0.019)	-	-0.231(0.293)	0.102(0.022)	3.610(4.057)
50	M1	0.072(0.121)	0.097(0.109)	0.019(0.032)	-0.027(0.212)	0.781(3.195)
	M2	0.013(0.209)	-	-0.210(0.197)	0.931(1.281)	2.926(3.566)
100	M1	0.017(0.039)	0.016(0.039)	-0.006(0.015)	-0.037(0.065)	0.549(2.348)
	M2	-0.035(0.106)	-	-0.220(0.209)	0.919(0.982)	2.595(2.814)
250	M1	0.003(0.011)	0.003(0.011)	0.001(0.005)	-0.036(0.021)	0.155(0.552)
	M2	-0.027(0.035)	-	-0.219(0.198)	0.558(0.341)	2.203(2.772)
500	M1	0.004(0.004)	0.006(0.004)	0.004(0.003)	-0.009(0.008)	0.009(0.195)
	M2	-0.024(0.015)	-	-0.218(0.194)	0.545(0.356)	2.066(2.578)

Table 5.6: RB and MSE values. True model = TM1, ICC = 0.88, β_2=0.3.

n	Model	$\beta_1(2.5)$	$\beta_2(0.3)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(5.28)$
20	M1	0.154(0.799)	-0.156(0.653)	-0.005(0.062)	0.034(0.596)	1.868(3.832)
	M2	0.077(0.510)	-	-0.005(0.058)	0.166(0.574)	1.554(2.975)
50	M1	0.059(0.140)	0.067(0.118)	-0.003(0.039)	-0.009(0.169)	0.882(2.504)
	M2	0.026(0.106)	-	-0.013(0.039)	0.181(0.186)	0.906(2.913)
100	M1	0.005(0.033)	-0.062(0.033)	0.005(0.013)	0.077(0.103)	0.373(1.651)
	M2	-0.019(0.0326)	-	-0.003(0.013)	0.247(0.162)	0.444(1.820)
250	M1	0.003(0.003)	-0.013(0.003)	0.006(0.007)	-0.0004(0.002)	0.027(0.395)
	M2	-0.003(0.003)	-	-0.004(0.007)	0.044(0.004)	0.172(0.495)
500	M1	0.003(0.002)	-0.011(0.002)	0.001(0.003)	0.003(0.004)	-0.014(0.241)
	M2	-0.010(0.004)	-	-0.011(0.005)	0.106(0.019)	0.086(0.289)

Results of Table 5.6 indicate that parameter estimators are not robust to model misspecification when the ICC is high. RB and MSE quantities decrease as sample size increases for both models. M1 and M2 have very close values for MSE of all parameters.

Table 5.7: RB and MSE values. True model = TM2, ICC = 0.20.

n	Model	$\beta_1(2.5)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.18)$
20	M2	0.042(0.135)	-0.009(0.071)	-0.021(0.025)	0.004(0.0001)
	M1	0.083(0.187)	-0.008(0.071)	-0.009(0.030)	0.005(0.0001)
50	M2	0.029(0.047)	0.005(0.031)	-0.0008(0.0105)	0.004(0.0001)
	M1	0.046(0.058)	0.006(0.031)	0.002(0.011)	0.005(0.0001)
100	M2	0.015(0.020)	0.006(0.014)	0.0008(0.004)	0.003(0.0002)
	M1	0.024(0.023)	0.007(0.014)	-0.007(0.005)	0.003(0.0002)
250	M2	0.003(0.008)	0.004(0.006)	0.004(0.002)	0.0004(0.0003)
	M1	0.007(0.009)	0.004(0.006)	0.003(0.002)	0.0008(0.0003)
500	M2	-0.0005(0.003)	0.003(0.003)	-0.003(0.0007)	0.002(0.0003)
	M1	0.001(0.003)	0.003(0.003)	-0.0009 (0.0008)	0.002(0.0003)

Table 5.8: RB and MSE values. True model = TM2, ICC = 0.50.

n	Model	$\beta_1(2.5)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.72)$
20	M2	0.081(0.167)	-0.0008(0.080)	-0.036(0.056)	0.122(0.085)
	M1	0.129(0.282)	0.001(0.081)	-0.039(0.058)	0.150(0.100)
50	M2	0.019(0.045)	0.019(0.023)	0.009(0.021)	0.062(0.054)
	M1	0.037(0.056)	0.020(0.023)	0.008(0.021)	0.072(0.058)
100	M2	0.017(0.026)	-0.0004(0.017)	-0.007(0.010)	0.047(0.037)
	M1	0.027(0.030)	-0.0005(0.017)	-0.007(0.012)	0.053(0.039)
250	M2	0.004(0.008)	0.005(0.005)	0.015(0.005)	0.026(0.009)
	M1	0.008(0.009)	0.005(0.005)	0.013(0.005)	0.028(0.009)
500	M2	0.0002(0.005)	0.0005(0.003)	0.007(0.002)	0.001(0.005)
	M1	0.0003(0.005)	0.0005(0.003)	0.008(0.002)	0.002(0.006)

Table 5.9: RB and MSE values. True model = TM2, ICC = 0.88.

n	Model	β_1(2.5)	κ(2)	μ_{b_0}(0)	$\sigma^2_{b_0}$(5.28)
20	M2	0.042(0.131)	0.008(0.081)	-0.014(0.175)	2.026(22.456)
	M1	0.083(0.174)	0.005(0.081)	-0.051(0.180)	2.340(26.319)
50	M2	0.018(0.066)	0.004(0.032)	0.003(0.104)	0.666(4.300)
	M1	0.037(0.079)	0.002(0.033)	-0.028(0.116)	0.748(4.643)
100	M2	0.016(0.026)	0.002(0.015)	0.011(0.054)	0.426(1.298)
	M1	0.023(0.031)	0.0002(0.017)	-0.005(0.069)	0.524(1.766)
250	M2	-0.004(0.002)	0.001(0.005)	0.008(0.002)	-0.015(0.561)
	M1	0.070(0.033)	-0.041(0.012)	-0.074(0.007)	0.575(1.140)
500	M2	0.0004(0.002)	0.005(0.003)	0.003(0.002)	0.017(0.226)
	M1	0.046(0.016)	-0.012(0.004)	-0.087(0.010)	0.333(0.415)

General results of Tables 5.7-5.9, in which true model is TM2, reveal that RB and MSE for all parameter estimators are smaller for M2, for correctly specified model, irrespective of the sample size and ICC values. This implies that the estimations of all parameters based on M2 are better than those for M1. Moreover, for all parameters, RB and MSE decrease for both models as the sample size increases from 20 to 500. Random effects variance and concentration parameter inferences are robust to model misspecification for low ICC. Parameter estimator for random intercept parameters are robust to model misspecification for mild ICC.

5.1.1.2 Interaction and Main Effect Mean Models

In this study, two types of true models (TM) are again considered. These are denoted by TM1 and TM2 and are as follows:

TM1: $\mu_{ij} = 2arctan(b_{0i} + \beta_1 x_{ij} + \beta_2 d_{1i} + \beta_3 d_{2i} + \beta_4 x_{ij} d_{1i} + \beta_5 x_{ij} d_{2i})$,

TM2: $\mu_{ij} = 2arctan(b_{0i} + \beta_1 x_{ij} + \beta_2 d_{1i} + \beta_3 d_{2i})$

As seen, TM1 is an interaction model whereas TM2 is a main effect model. We also considered two different values for coefficients of interaction effects. β_4 and β_5 are set at (2, 2.5) or (0.3, 0.5) representing effective and relatively ineffective interaction terms to control the simulations for interaction term effect size. This simulation study is controlled for four different sample sizes n=20, 50, 100, 250. Dichotomous dummy variables, d_1 and d_2 are used to include a time-independent categorical variable with three levels. In order to generate this categorical variable, the inverse transformation method (Martinez et. al., 2002) is used with the probabilities 0.33, 0.34, 0.33.

Fitted models denoted by M1 and M2 are as follows

M1: $\mu_{ij} = 2\arctan(b_{0i} + \beta_1 x_{ij} + \beta_2 d_{1i} + \beta_3 d_{2i} + \beta_4 x_{ij} d_{1i} + \beta_5 x_{ij} d_{2i}),$

M2: $\mu_{ij} = 2\arctan(b_{0i} + \beta_1 x_{ij} + \beta_2 d_{1i} + \beta_3 d_{2i}).$

When data are generated from TM1, RB and MSEs of Bayesian estimators of the parameters of interest under different ICC, β_4 and β_5 are presented in Tables 5.10-5.15. In these tables, Model=M1 lines determine the behaviour of the estimators when fitted model correctly specifies the underlying true model whereas Model=M2 lines determine those under mean model misspecification. For Tables 5.16-5.18, data are generated from TM2. Model=M2 lines indicate the behaviour of the estimators for fitting true model and Model=M1 lines denote the results when the fitted model is misspecified.

According to general results of Tables 5.10, 5.12 and 5.14, when the sizes of interaction effects are large, RB and MSE for all parameters are general smaller for M1 (the correct model). As the sample size increases from 20 to 250, these measures decrease when the model is correctly specified. Parameter estimators are not robust to model misspecification.

Table 5.10: RB and MSE values. TM = TM1, ICC = 0.20, $\beta_4 = 2$, $\beta_5 = 2.5$.

n	Model	β_1(2.5)	β_2(1.5)	β_3(1.8)	β_4(2)	β_5(2.5)	κ(2)	μ_{b_0}(0)	$\sigma^2_{b_0}$(0.18)
20	M1	0.204(1.112)	0.359(1.087)	0.263(0.903)	0.587(7.066)	0.541(6.973)	0.004(0.059)	0.019(0.144)	0.019(0.000)
	M2	0.696(3.478)	-0.045(0.467)	-0.184(0.490)	-	-	-0.031(0.052)	0.044(0.310)	0.020(0.000)
50	M1	0.049(0.161)	0.154(0.266)	0.126(0.257)	0.200(1.336)	0.217(1.561)	0.007(0.025)	-0.042(0.026)	0.036(0.0001)
	M2	0.598(2.412)	-0.045(0.129)	-0.164(0.215)	-	-	-0.038(0.029)	-0.074(0.068)	0.045(0.0003)
100	M1	0.041(0.059)	0.023(0.062)	0.066(0.112)	0.001(0.287)	0.065(0.534)	0.002(0.013)	-0.010(0.015)	0.025(0.0002)
	M2	0.540(1.884)	-0.126(0.097)	-0.198(0.178)	-	-	-0.037(0.016)	-0.017(0.033)	0.037(0.0002)
250	M1	0.019(0.022)	0.023(0.023)	0.024(0.034)	0.030(0.116)	0.042(0.180)	-0.0001(0.006)	-0.002(0.005)	0.010(0.0003)
	M2	0.534(1.816)	-0.137(0.064)	-0.222(0.186)	-	-	-0.045(0.014)	-0.002(0.013)	0.031(0.0003)

Table 5.11: RB and MSE values. TM = TM1, ICC = 0.20, $\beta_4 = 0.3$, $\beta_5 = 0.5$.

n	Model	β_1(2.5)	β_2(1.5)	β_3(1.8)	β_4(0.3)	β_5(0.5)	κ(2)	μ_{b_0}(0)	$\sigma^2_{b_0}$(0.18)
20	M1	0.141(0.382)	0.132(0.362)	0.195(0.451)	0.686(0.750)	-0.667(1.135)	0.024(0.074)	0.053(0.082)	0.030(0.000)
	M2	0.199(0.394)	0.027(0.202)	0.006(0.203)	-	-	0.023(0.070)	0.045(0.088)	0.031(0.000)
50	M1	0.108(0.204)	0.062(0.133)	0.073(0.156)	-0.190(0.337)	-0.812(0.350)	0.015(0.031)	0.029(0.034)	0.021(0.0001)
	M2	0.148(0.193)	0.013(0.075)	-0.040(0.088)	-	-	0.013(0.030)	0.030(0.035)	0.021(0.0001)
100	M1	0.035(0.088)	0.050(0.050)	0.039(0.060)	0.194(0.132)	-0.776(0.171)	0.013(0.021)	-0.004(0.012)	0.021(0.0002)
	M2	0.130(0.135)	0.011(0.033)	-0.053(0.040)	-	-	0.010(0.021)	-0.004(0.014)	0.022(0.0002)
250	M1	0.007(0.018)	0.020(0.019)	0.009(0.020)	0.180(0.053)	-0.790(0.062)	0.000(0.004)	0.002(0.004)	-0.002(0.0002)
	M2	0.110(0.085)	-0.009(0.013)	-0.069(0.029)	-	-	-0.003(0.004)	0.002(0.006)	0.0003(0.0002)

Table 5.12: RB and MSE values. TM = TM1, ICC = 0.50, $\beta_4 = 2$, $\beta_5 = 2.5$.

n	Model	β_1(2.5)	β_2(1.5)	β_3(1.8)	β_4(2)	β_5(2.5)	κ(2)	μ_{b_0}(0)	$\sigma^2_{b_0}$(0.72)
20	M1	0.017(0.017)	-0.019(0.017)	-0.012(0.017)	0.008(0.007)	0.009(0.005)	0.050(0.094)	-0.005(0.022)	0.081(0.055)
	M2	0.212(0.300)	-0.083(0.028)	-0.099(0.048)	-	-	-0.022(0.073)	-0.163(0.047)	0.033(0.053)
50	M1	0.006(0.021)	0.003(0.028)	0.003(0.019)	-0.0001(0.011)	-0.006(0.011)	0.009(0.024)	-0.002(0.018)	0.048(0.048)
	M2	0.281(0.523)	-0.086(0.037)	-0.129(0.073)	-	-	-0.035(0.029)	-0.156(0.046)	0.004(0.069)
100	M1	0.012(0.018)	0.005(0.020)	0.008(0.023)	-0.002(0.017)	-0.003(0.014)	-0.004(0.015)	-0.002(0.013)	-0.001(0.027)
	M2	0.364(0.851)	-0.093(0.037)	-0.154(0.101)	-	-	-0.047(0.025)	-0.127(0.035)	-0.0547(0.033)
250	M1	-0.0008(0.002)	-0.005(0.002)	0.002(0.001)	0.001(0.0008)	-0.0004(0.0008)	0.003(0.005)	0.004(0.002)	0.036(0.015)
	M2	0.212(0.284)	-0.078(0.015)	-0.106(0.038)	-	-	-0.057(0.019)	-0.169(0.031)	-0.068(0.018)

Table 5.13: RB and MSE values. TM = TM1, ICC = 0.50, $\beta_4 = 0.3$, $\beta_5 = 0.5$.

n	Model	β_1(2.5)	β_2(1.5)	β_3(1.8)	β_4(0.3)	β_5(0.5)	κ(2)	μ_{b_0}(0)	$\sigma^2_{b_0}$(0.72)
20	M1	0.014(0.022)	-0.009(0.019)	0.003(0.015)	0.056(0.012)	-0.797(0.013)	0.031(0.109)	0.036(0.023)	0.132(0.094)
	M2	0.064(0.049)	-0.017(0.021)	-0.023(0.017)	-	-	0.026(0.108)	-0.002(0.022)	0.096(0.081)
50	M1	0.014(0.020)	-0.007(0.018)	-0.013(0.017)	0.044(0.016)	-0.795(0.018)	-0.008(0.026)	0.024(0.017)	0.073(0.056)
	M2	0.084(0.065)	-0.016(0.019)	-0.055(0.028)	-	-	-0.013(0.026)	-0.011(0.017)	0.042(0.051)
100	M1	0.011(0.012)	0.006(0.019)	-0.008(0.021)	-0.063(0.025)	-0.795(0.018)	0.0004(0.012)	0.006(0.011)	0.047(0.023)
	M2	0.087(0.059)	0.002(0.018)	-0.060(0.033)	-	-	-0.003(0.012)	-0.019(0.012)	0.031(0.021)
250	M1	-0.001(0.007)	-0.010(0.013)	0.008(0.012)	0.046(0.018)	-0.788(0.016)	-0.003(0.004)	-0.003(0.006)	0.0004(0.008)
	M2	0.085(0.053)	-0.018(0.013)	-0.047(0.020)	-	-	-0.007(0.005)	-0.029(0.009)	-0.008(0.009)

Table 5.14: RB and MSE values. TM = TM1, ICC = 0.88, $\beta_4 = 2$, $\beta_5 = 2.5$.

n	Model	β_1(2.5)	β_2(1.5)	β_3(1.8)	β_4(2)	β_5(2.5)	κ(2)	μ_{b_0}(0)	$\sigma^2_{b_0}$(5.28)
20	M1	0.0001(0.0003)	0.0003(0.0001)	-0.0007(0.0001)	-0.0006(0.0001)	0.0004(0.0001)	-0.011(0.100)	-0.001(0.0002)	0.335(5.888)
	M2	0.019(0.003)	-0.002(0.0001)	-0.002(0.0001)	-	-	-0.077(0.102)	-0.009(0.0003)	0.961(5.412)
50	M1	-0.0004(0.0007)	-0.001(0.0002)	-0.0001(0.0002)	-0.0005(0.0002)	0.0001(0.0001)	0.022(0.047)	-0.002(0.0006)	0.109(3.758)
	M2	0.049(0.016)	-0.007(0.0003)	-0.007(0.0004)	-	-	-0.048(0.040)	-0.020(0.001)	0.050(5.731)
100	M1	0.0004(0.001)	0.0004(0.0005)	0.0015(0.0003)	0.001(0.0003)	-0.001(0.0002)	0.009(0.022)	0.001(0.0008)	0.058(1.203)
	M2	0.087(0.049)	-0.008(0.0006)	-0.009(0.0006)	-	-	-0.052(0.028)	-0.031(0.002)	-0.047(1.759)
250	M1	0.001(0.002)	-0.001(0.0005)	-0.002(0.0007)	-0.0004(0.0007)	0.0004(0.0006)	-0.002(0.005)	-0.001(0.001)	0.008(0.486)
	M2	0.166(0.174)	-0.019(0.002)	-0.024(0.003)	-	-	-0.052(0.016)	-0.058(0.006)	-0.093(0.945)

Table 5.15: RB and MSE values. TM = TM1, ICC = 0.88, $\beta_4 = 0.3$, $\beta_5 = 0.5$.

n	Model	β_1(2.5)	β_2(1.5)	β_3(1.8)	β_4(0.3)	β_5(0.5)	κ(2)	μ_{b_0}(0)	$\sigma^2_{b_0}$(5.28)
20	M1	0.0006(0.0004)	0.0006(0.0001)	0.0005(0.0001)	0.003(0.0002)	-0.799(0.0002)	0.001(0.084)	0.003(0.0003)	0.215(9.648)
	M2	0.006(0.0008)	0.0003(0.0001)	-0.0006(0.0001)	-	-	-0.007(0.081)	0.001(0.0003)	0.146(9.247)
50	M1	-0.0005(0.001)	0.0002(0.0002)	0.0004(0.0002)	-0.003(0.0005)	-0.799(0.0004)	-0.006(0.033)	0.003(0.0006)	0.010(2.866)
	M2	0.013(0.002)	-0.002(0.0002)	-0.002(0.0002)	-	-	-0.013(0.033)	-0.002(0.0006)	-0.064(2.713)
100	M1	0.001(0.002)	-0.0009(0.0004)	0.0005(0.0004)	-0.003(0.0007)	-0.800(0.0006)	0.001(0.015)	0.002(0.001)	0.039(1.494)
	M2	0.024(0.006)	-0.002(0.0004)	-0.003(0.0004)	-	-	-0.003(0.015)	-0.006(0.001)	-0.029(1.355)
250	M1	0.003(0.001)	0.002(0.0008)	-0.001(0.0007)	0.009(0.001)	-0.798(0.001)	0.003(0.007)	0.006(0.001)	0.005(0.629)
	M2	0.045(0.015)	-0.0005(0.0009)	-0.008(0.001)	-	-	-0.0009(0.007)	-0.009(0.002)	-0.046(0.645)

Results of Tables 5.11, 5.13 and 5.15 indicate that when the effects of interaction terms are small, RB and MSE for random intercept parameters are very close to each other for both models irrespective of the sample size when ICC is low (Table 5.11). As sample size increases, RB and MSE for all parameter estimators decrease when ICC is low, when fitted model is correctly specified. Parameter estimators are not robust to model misspecification for all ICC.

According to the results of Table 5.16, regression coefficients have smaller RB and MSE values when the model is correctly specified. When random intercept parameters are concerned, RB and MSE are very close to each other for both models. Random effects parameters inferences are robust to model misspecification. As the sample size increases, RB and MSE for all parameters of interest decrease for both models and both RB and MSE of all parameters are at an acceptable level for all sample sizes.

Table 5.16: RB and MSE values. True model = TM2, ICC = 0.20.

n	Model	$\beta_1(2.5)$	$\beta_2(1.5)$	$\beta_3(1.8)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.18)$
20	M2	0.088(0.198)	0.198(0.282)	0.142(0.659)	0.015(0.094)	-0.040(0.081)	0.030(0.0001)
	M1	0.177(0.769)	0.336(1.105)	0.292(1.232)	0.010(0.094)	-0.042(0.105)	0.030(0.0001)
50	M2	0.044(0.058)	0.049(0.084)	0.049(0.081)	-0.002(0.025)	0.003(0.025)	0.025(0.0001)
	M1	0.031(0.111)	0.096(0.146)	0.111(0.196)	-0.003(0.025)	0.006(0.025)	0.026(0.0001)
100	M2	0.013(0.025)	0.016(0.029)	0.012(0.029)	0.007(0.019)	-0.010(0.011)	0.021(0.0002)
	M1	0.018(0.054)	0.027(0.037)	0.040(0.066)	0.007(0.019)	-0.012(0.012)	0.024(0.0002)
250	M2	0.005(0.007)	-0.003(0.010)	-0.0001(0.013)	0.0009(0.004)	-0.01(0.004)	-0.017(0.0002)
	M1	0.008(0.019)	0.003(0.013)	0.008(0.024)	0.001(0.04)	0.009(0.004)	-0.017(0.0002)

Table 5.17: RB and MSE values. True model = TM2, ICC = 0.50.

n	Model	$\beta_1(2.5)$	$\beta_2(1.5)$	$\beta_3(1.8)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.72)$
20	M2	0.005(0.0154)	0.004(0.019)	0.001(0.018)	0.014(0.080)	-0.007(0.024)	0.123(0.086)
	M1	-0.126(0.129)	0.070(0.029)	0.112(0.057)	-0.064(0.075)	0.201(0.066)	0.741(0.566)
50	M2	0.014(0.024)	-0.008(0.014)	0.003(0.021)	0.010(0.021)	0.016(0.017)	0.097(0.043)
	M1	-0.157(0.188)	0.099(0.039)	0.162(0.108)	-0.043(0.030)	0.214(0.061)	0.509(0.233)
100	M2	-0.0003(0.002)	0.002(0.001)	0.001(0.001)	-0.011(0.014)	-0.003(0.002)	-0.006(0.018)
	M1	-0.071(0.034)	0.052(0.007)	0.064(0.014)	-0.092(0.0483)	0.160(0.027)	1.119(0.752)
250	M2	-0.002(0.003)	0.003(0.002)	0.002(0.002)	0.003(0.008)	-0.002(0.002)	0.013(0.010)
	M1	-0.132(0.112)	0.0890(0.020)	0.122(0.050)	-0.067(0.025)	0.226(0.053)	0.766(0.339)

Table 5.18: RB and MSE values. True model = TM2, ICC = 0.88.

n	Model	$\beta_1(2.5)$	$\beta_2(1.5)$	$\beta_3(1.8)$	$\kappa(2)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(5.28)$
20	M2	0.0005(0.0006)	-0.002(0.0001)	-0.0009(0.0001)	0.017(0.010)	-0.005(0.0003)	0.321(2.471)
	M1	-0.009(0.0008)	0.0006(0.0001)	0.002(0.0001)	-0.076(0.114)	0.004(0.0002)	1.138(5.974)
50	M2	-0.002(0.001)	-0.0004(0.0003)	-0.0006(0.0002)	0.009(0.027)	0.001(0.0007)	0.037(2.211)
	M1	-0.023(0.004)	0.005(0.0001)	0.005(0.0002)	-0.079(0.053)	0.018(0.0006)	0.903(3.679)
100	M2	0.004(0.002)	-0.002(0.0003)	0.0001(0.0004)	0.002(0.021)	-0.003(0.0007)	0.010(1.972)
	M1	-0.035(0.009)	0.007(0.0003)	0.010(0.0006)	-0.079(0.044)	0.027(0.001)	0.907(2.677)
250	M2	0.004(0.003)	-0.002(0.001)	0.001(0.0006)	0.005(0.007)	0.003(0.002)	0.021(0.502)
	M1	-0.081(0.043)	0.016(0.001)	0.023(0.002)	-0.064(0.023)	-0.065(0.056)	0.669(1.109)

Table 5.17 and Table 5.18 show that M2 has smaller RB and MSE for all parameters irrespective of the sample size when mild and high ICC are concerned. This means that, all parameters of interest are better estimated based on M2, as expected. As sample size increases, RB and MSE for all parameters converge to 0 when the model is correctly specified. Parameter estimators are not robust to model misspecification.

5.1.2 Different Distributions

In this section, an interesting simulation study is designed in order to evaluate the performance of the vM based parameter estimators when the true data generation process is from different circular distributions. Data sets are generated from WC distribution and vM distribution, then vM distribution is fitted to each of data generated from the distributions mentioned earlier. This scheme is repeated 100 times. The reasons for choosing WC distribution are as follows: *i.* WC distribution is symmetric and unimodal distribution, like vM distribution, *ii.* for analyzing symmetric unimodal circular data, WC distribution is proposed as an alternative to vM distribution for suitable the dispersion parameters (Kent and Tyler, 1988). We generate the WC distributed circular responses using the following

$$\begin{aligned} &\theta_{ij} \sim WC(\mu_{ij}, \rho),\ i = 1, ...n,\ j = 1, ..., m, \\ &\mu_{ij} = (b_{0i} + \boldsymbol{\beta X}_{ij})\ [mod\, 2\pi], \end{aligned} \tag{5.5}$$

where μ_{ij} and ρ denote the mean direction and the mean resultant length, respectively. b_{0i} is subject specific random intercept, $\boldsymbol{\beta}$ is a vector of regression coefficients and X_i is a vector of linear covariates. For the circular regression purpose, the range of μ_{ij} is converted from (0, 2π) to ($-\pi$, π). Only one linear covariate following a standard normal distribution is used. This simulation study is controlled for four different sample sizes, n=50, 100, 250, 500 and three different ICC values, ICC=0.20, 0.50, 0.88.

True models denoted by TM1 and TM2 are given in the hierarchical representation as follows

TM1: $\theta_{ij} \sim vM(\mu_{ij}, \kappa)$
$\mu_{ij} = 2\arctan(b_{0i} + \beta_1 x_{ij})$,

TM2: $\theta_{ij} \sim WC(\mu_{ij}, \rho)$
$\mu_{ij} = (b_{0i} + \beta_1 x_{ij})\ [mod\, 2\pi]$

where κ and ρ are set at 2 and 0.64 respectively so that the two distributions have the same circular dispersion, which is 0.72. Fitted model is TM1. RB and MSEs of Bayesian estimators of the parameters under different ICC and sample size are presented in Tables 5.19-5.21. In these tables, results corresponding to True Model=TM1 are the results when there is no misspecification whereas results corresponding to True Model=TM2 are the results under misspecification.

According to the results of Tables 5.19-5.21 RB and MSE of parameter estimators are generally smaller for TM1. As the sample size increases, these quantities decrease for TM1 irrespective of ICC values. Parameter estimations are better based on TM1 for all ICC values. Parameter inferences are not robust to distribution misspecification.

Table 5.19: RB and MSE values. ICC = 0.20.

n	True Model	$\beta_1(2.5)$	$\delta(0.72)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.18)$
50	TM1	0.023(0.039)	0.013(0.008)	-0.002(0.008)	0.018(0.0001)
	TM2	-0.055(0.060)	-0.949(0.583)	0.007(0.0091)	-0.047(0.0001)
100	TM1	0.016(0.019)	-0.008(0.003)	0.005(0.005)	0.028(0.0002)
	TM2	-0.091(0.069)	-0.959(0.566)	-0.010(0.004)	-0.106(0.0004)
250	TM1	0.009(0.008)	-0.010(0.001)	-0.001(0.001)	0.008(0.0002)
	TM2	-0.103(0.073)	-0.963(0.559)	0.005(0.002)	-0.224(0.002)
500	TM1	0.003(0.003)	-0.0003(0.0007)	0.001(0.001)	-0.014(0.0003)
	TM2	-0.110(0.079)	-0.961(0.560)	-0.005(0.001)	-0.331(0.004)

Table 5.20: RB and MSE values. ICC = 0.50.

n	True Model	$\beta_1(2.5)$	$\delta(0.72)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(0.72)$
50	TM1	0.021(0.052)	0.003(0.006)	0.003(0.018)	0.094(0.041)
	TM2	0.030(0.056)	-0.847(0.409)	0.017(0.027)	-0.305(0.061)
100	M1	0.006(0.033)	0.008(0.004)	-0.012(0.010)	0.058(0.032)
	M2	-0.067(0.063)	-0.793(0.344)	0.0004(0.010)	-0.417(0.098)
250	TM1	-0.0005(0.008)	-0.001(0.001)	-0.006(0.004)	0.018(0.010)
	TM2	-0.096(0.071)	-0.815(0.352)	-0.006(0.004)	-0.505(0.137)
500	TM1	0.007(0.004)	-0.001(0.0007)	0.001(0.002)	0.028(0.007)
	TM2	-0.112(0.085)	-0.811(0.345)	0.003(0.002)	-0.562(0.166)

Table 5.21: RB and MSE values. ICC = 0.88.

n	True Model	$\beta_1(2.5)$	$\delta(0.72)$	$\mu_{b_0}(0)$	$\sigma^2_{b_0}(5.28)$
50	TM1	0.033(0.064)	0.002(0.009)	-0.071(0.126)	0.043(1.658)
	TM2	-0.106(0.131)	-0.036(0.815)	0.055(0.136)	-0.054(0.935)
100	TM1	0.009(0.029)	-0.001(0.004)	0.005(0.078)	0.095(1.593)
	TM2	-0.137(0.143)	0.443(2.132)	0.020(0.130)	-0.131(1.227)
250	TM1	0.011(0.014)	-0.0004(0.001)	0.014(0.027)	0.026(0.518)
	TM2	-0.182(0.221)	-0.408(0.141)	-0.003(0.049)	-0.280(2.532)
500	TM1	0.002(0.006)	0.0004(0.0008)	0.006(0.012)	-0.001(0.261)
	TM2	-0.183(0.216)	-0.388(0.109)	0.019(0.032)	-0.305(2.784)

5.2 Model Assessment, Comparison and Selection

In this section, model assessment, comparison and selection criteria detailed in Chapter 4 are evaluated for their ability to select the true model. This simulation scenario is divided into groups as in Section 5.1.

5.2.1 Selection Over the Mean Models

The performances of criteria are evaluated for linear-quadratic mean models (Section 5.2.1.1) and main-interaction mean models (Section 5.2.1.2). The frequency of selecting the true model is used to assess the performances of criteria.

5.2.1.1 Quadratic and Linear Mean Models

Tables 5.22-5.24 show the frequency of selecting the true model (out of 100 replications) for each criterion when the true model is TM1 mentioned in Section 5.1.1.1. When the true model is TM2 mentioned in Section 5.1.1.1, the frequency of selecting the true model for each criterion is presented in Table 5.25.

According to Tables 5.22-5.24, when the effect of quadratic term is large ($\beta_2 = 1.5$), in other words, when there is a relatively emphasized nonlinearity in data generation model, the criteria perform equivalently. When the effect of quadratic term is small

($\beta_2 = 0.3$) the frequencies of selecting true model for CPD_1 and CPD_2 are greater than those of $PAPE_1$ and $PAPE_2$ for all ICC and sample sizes. The frequency of selecting true model increases for all criteria as the sample size increases for all ICC values when the quadratic effect is relatively ineffective.

Table 5.22: Frequency of selecting the true model, low ICC (0.20).

β_2	0.3					1.5				
n	20	50	100	250	500	20	50	100	250	500
CPD_1	49	61	77	96	98	100	100	100	100	100
$PAPE_1$	32	55	60	86	90	97	100	100	100	100
CPD_2	49	60	76	96	98	100	100	100	100	100
$PAPE_2$	35	56	66	87	92	97	100	100	100	100

Table 5.23: Frequency of selecting the true model, mild ICC (0.50).

β_2	0.3					1.5				
n	20	50	100	250	500	20	50	100	250	500
CPD_1	49	68	76	90	98	100	100	100	100	100
$PAPE_1$	36	56	58	79	83	99	100	100	100	100
CPD_2	49	68	76	91	98	100	100	100	100	100
$PAPE_2$	34	56	57	79	83	99	100	100	100	100

Table 5.24: Frequency of selecting the true model, high ICC (0.88).

β_2	0.3					1.5				
n	20	50	100	250	500	20	50	100	250	500
CPD_1	36	71	83	84	99	98	100	100	100	100
$PAPE_1$	23	50	70	75	97	98	100	100	100	100
CPD_2	36	73	83	84	99	98	100	100	100	100
$PAPE_2$	23	52	67	75	97	98	100	100	100	100

According to Table 5.25, when the true model is the linear model, performances of the criteria improve with larger ICC. PAPE type criteria perform in general better than CPD type criteria except when ICC is low and sample size is large. Between the $PAPE_1$ and $PAPE_2$, $PAPE_2$ performs generally better than $PAPE_1$. It appears that CPD_1 and CPD_2 have tendency to select the more complex model as the best fitting model. On the other hand, $PAPE_1$ or $PAPE_2$ seem to better capture the difference between the simpler and the more complex model. The performance of $PAPE_1$ and $PAPE_2$ is prospering and these criteria seem to be promising.

Table 5.25: Frequency of selecting the true model. Data generation model: Linear Model.

ICC	0.20					0.50					0.88				
n	20	50	100	250	500	20	50	100	250	500	20	50	100	250	500
CPD_1	62	66	60	64	51	63	63	59	59	57	70	78	65	100	100
$PAPE_1$	70	64	50	54	47	73	77	73	67	60	81	83	78	100	100
CPD_2	61	65	60	64	52	63	63	59	57	57	70	78	65	100	100
$PAPE_2$	68	67	45	59	46	78	77	73	67	62	80	87	81	100	100

5.2.1.2 Interaction and Main Effect Mean Models

When the true model is TM1 mentioned in Section 5.1.1.2 under different ICC, sample size and the size of interaction effects, the frequency of selecting the true model for each criterion is presented in Tables 5.26-5.28. Table 5.29 indicates the frequency of selecting the true model for each criterion, when the true model is TM2 mentioned in Section 5.1.1.2 under different ICC and sample size.

Table 5.26: Frequency of selecting the true model, ICC=0.20

(β_4,β_5)	(0.3, 0.5)				(2, 2.5)			
n	20	50	100	250	20	50	100	250
CPD_1	39	40	60	77	79	100	100	100
$PAPE_1$	14	11	12	15	32	55	68	87
CPD_2	38	40	60	78	79	100	100	100
$PAPE_2$	12	10	12	13	28	52	66	84

Table 5.27: Frequency of selecting the true model, ICC=0.50

β_4,β_5	(0.3, 0.5)				(2, 2.5)			
n	20	50	100	250	20	50	100	250
CPD_1	62	74	71	87	92	96	99	100
$PAPE_1$	21	13	18	11	24	17	18	4
CPD_2	60	74	70	87	92	96	99	100
$PAPE_2$	23	10	15	7	23	15	16	2

Table 5.28: Frequency of selecting the true model, ICC=0.88

β_4,β_5	(0.3, 0.5)				(2, 2.5)			
n	20	50	100	250	20	50	100	250
CPD_1	72	72	75	89	82	99	100	100
$PAPE_1$	22	15	5	1	41	30	20	10
CPD_2	70	72	75	89	83	99	100	100
$PAPE_2$	19	13	4	1	37	27	17	8

As seen from the results of Tables 5.26-5.28, the performances of CPD_1 and CPD_2 are better than those of $PAPE_1$ and $PAPE_2$ irrespective of sample size and ICC value.

According to results of Table 5.29, in which data are generated from TM2, when ICC is low, there is an interesting picture of the frequency of selecting true model for each criterion. $PAPE_1$ and $PAPE_2$ outperform CPD_1 and CPD_2 irrespective of the sample size for low ICC, especially for the relatively small sample size. $PAPE_1$ and $PAPE_2$ again outperform CPD_1 and CPD_2 for both mild and high ICC for the relatively small sample size. Moreover, the performances of all criteria are more satisfactory for both mild and high ICC. As the sample size increases, the performances of all criteria get better for mild and high ICC. In order to select the best fitted model when data sets are generated from TM2, $PAPE_1$ and $PAPE_2$ can be preferred irrespective of the sample size and ICC, especially for the relatively small sample size.

Table 5.29: Frequency of selecting the true model. Data generation model: Main Effect Model.

ICC	0.20				0.50				0.88			
n	20	50	100	250	20	50	100	250	20	50	100	250
CPD_1	66	64	53	63	90	96	100	100	95	98	100	100
$PAPE_1$	86	89	79	77	99	100	100	100	99	100	100	100
CPD_2	66	64	55	63	90	96	100	100	95	98	100	100
$PAPE_2$	86	90	79	75	99	100	100	100	99	100	100	100

5.2.2 Different Distributions

In this section, the results of Monte Carlo simulation study are presented to compare the performance of criteria over 100 Monte Carlo replications, when data sets are generated from two different distributions. The logic behind data generation process is the same as in Section 5.1.2. We introduce a new statistic called Average Stability Coefficient (ASC) to compare the performances of the model selection criteria in the settings such as the one presented in this section. A brief description of ASC is presented below. In this simulation study, we consider two different data generation processes where two different circular distributions are used for each data generation process. On the other hand, that in circular data literature, there is only one circular regression model, namely vM regression. Hence, the frequency of selecting true model can not be calculated for this simulation study and a new quantity should be introduced (e.g our average stability coefficient). It is straightforward that when the data sets are generated from WC distribution (TM2 mentioned in Section 5.1.2) and the fitted distribution is vM distribution, each criterion is supposed to take maximum value. On the other hand, when data are generated from vM distribution (TM1 mentioned in Section 5.1.2) since the distribution used in both processes is the same, each criterion is supposed to take minimum value. The difference between values of each criterion for TM1 and TM2 should be as large as possible to have a good performance in selecting the best fitting model. This difference can be measured by using the ratio of value of criterion for TM2 to that for TM1 in each Monte Carlo replication. The ASC is proposed as follows;

$$ASC_{tm} = \frac{1}{M}\sum_{i=1}^{M} tm_2/tm_1, \tag{5.6}$$

where M is the number of Monte Carlo replications, tm denotes a particular criterion, tm_1 denotes the value of criterion of interest when the data set is generated from TM1, and tm_2 denotes the criterion value when TM2 is used to generate data set for each Monte Carlo replication. For instance, in order to calculate ASC for CPD_1, the value of this criterion for TM2 is divided by its value for TM1 for each Monte Carlo replication and then, the average of these M ratios is found to obtain ASC for this criterion. This coefficient shows the stability of the model selection tool on a particular selection and the further it is upwardly away from the unity, the more strongly the criterion selects the true model. This coefficient allows us to examine the decisiveness of the method. Tables 5.30-5.32 indicate ASC of each criterion for low, mild and high ICC, respectively.

According to the results of Tables 5.30-5.32 ASC values of CPD_2 are greater than those of other criteria irrespective of the sample size and ICC. This means that CPD_2 can select more strongly the true model and it outperforms the others. As ICC increases, ASC of CPD_1 and CPD_2 consistently increases. There may be a relationship between ICC and the decisiveness of the method on selection of the true model.

Table 5.30: ASC values over 100 Monte Carlo replications, ICC=0.20

n	50	100	250	500
$ASC(CPD_1)$	1.358	1.375	1.377	1.372
$ASC(PAPE_1)$	1.230	1.233	1.244	1.242
$ASC(CPD_2)$	1.518	1.544	1.547	1.538
$ASC(PAPE_2)$	1.328	1.334	1.348	1.344

Table 5.31: ASC values over 100 Monte Carlo replications, ICC=0.50

n	50	100	250	500
$ASC(CPD_1)$	1.414	1.428	1.430	1.432
$ASC(PAPE_1)$	1.268	1.281	1.287	1.285
$ASC(CPD_2)$	1.598	1.688	1.622	1.626
$ASC(PAPE_2)$	1.368	1.387	1.397	1.391

Table 5.32: ASC values over 100 Monte Carlo replications, ICC=0.88

n	50	100	250	500
$ASC(CPD_1)$	1.494	1.512	1.487	1.491
$ASC(PAPE_1)$	1.182	1.183	1.171	1.181
$ASC(CPD_2)$	1.715	1.741	1.705	1.710
$ASC(PAPE_2)$	1.233	1.235	1.220	1.233

The computational times of the simplest case and the most complex case for one Monte Carlo iteration are equal to 628.05 and 4223.05 seconds, respectively. In the simplest case, data sets are generated from a linear mean model, sample size is equal to 20 and ICC is low. In the most complex case, data sets are generated from an interaction mean model with relatively ineffective interaction effects, sample size is equal to 250 and ICC is high. MC study was repeated 250 times for the simplest case and it was seen that there was no significant difference between results of MS study with 100 repeated and results of MC study with 250 repeated. For more complex case, MC studies would take longer time. Therefore, the number of MC iteration is selected as 100.

5.3 An Additional Simulation Study

Some authors, e.g. Lagona (2016), include an intercept parameter defined on the unit circle instead of a linear intercept parameter that is defined in the link function, in circular regression models. In the first approach, intercept is the initial circular response (θ_{ij}) whereas in the latter model, it is the initial value of $\tan(\theta_{ij}/2)$. These two approaches have never been compared in the literature.

In this simulation study, we provide the comparison for performances of criteria under the first approach. Interested reader further can compare the results in this section and those in the previous section under the latter approach, which is our main model of interest.

We found out that, in the longitudinal setting, first approach has a considerable computational burden. Therefore we do not pursuit further to simulation scenarios other than the ones presented here.

5.3.1 Selection Over the Mean Models

The performances of criteria are evaluated for quadratic mean models (Section 5.3.1.1) and interaction mean models (Section 5.3.1.2). The frequency of selecting the true model is used to assess the performances of criteria.

5.3.1.1 Quadratic Mean Models

In this study, one type of true model (TM), in which data are generated from a quadratic mean model is considered and is as follows:

$$\text{TM: } \mu_{ij} = b_{0i} + 2\arctan(\beta_1 x_{ij} + \beta_2 x_{ij}^2),$$

where b_{0i} is random intercept for subject *i* and it follows a truncated normal distribution lying within the interval (-π, π). Lagona (2016) also used a circular regression model with an offset parameter lying within the interval (-π, π), as in the TM. This study is controlled for two different sample sizes, n=500, 1000 and ICC values, ICC=0.33, 0.50. We also considered two different values for quadratic term. β_2 is set at 1.5 or 0.3 representing effective and relatively ineffective quadratic term to control the simulations for quadratic term effect size. Just single time-dependent linear explanatory variable x_{ij} is used and produced by random draws from a standard normal distribution.

Fitted models are labeled by M1 and M2 and are as follows

$$\text{M1: } \mu_{ij} = b_{0i} + 2\arctan(\beta_1 x_{ij} + \beta_2 x_{ij}^2)$$

$$\text{M2: } \mu_{ij} = b_{0i} + 2\arctan(\beta_1 x_{ij}).$$

Table 5.33 shows the frequency of selecting the true model under different sample sizes, ICC values and sizes of quadratic term. According to results of Table 5.33, for both low and mild ICC, all criteria successfully select the correct model when candidate set includes models with same distributions with same concentrations but different mean models with linear and quadratic term, when the effect of quadratic term is relatively emphasized ($\beta_2 = 1.5$). When the effect of quadratic term is small, $PAPE_1$ and $PAPE_2$ outperform the other criteria irrespective of sample size and ICC values. $PAPE_1$ in particular select the correct model more strongly even when the true effect size of quadratic term is small.

Table 5.33: Frequency of selecting the true model.

ICC	0.33				0.50		
β_2	0.3		1.5		0.3	1.5	
n	500	1000	500	1000	500	500	1000
CPD_1	60	55	100	100	58	100	100
$PAPE_1$	64	66	100	100	63	100	100
CPD_2	60	55	100	100	58	100	100
$PAPE_2$	63	66	100	100	61	100	100

5.3.1.2 Interaction Mean Models

In this study, one type of true model (TM) in which data are generated from a interaction mean model is considered and is as follows:

$$\text{TM: } \mu_{ij} = b_{0i} + 2\arctan(\beta_1 x_{ij} + \beta_2 d_{1i} + \beta_3 d_{2i} + \beta_4 x_{ij} d_{1i} + \beta_5 x_{ij} d_{2i}),$$

We also considered two different values for coefficients of interaction effects. β_4 and β_5 are set at (2, 2.5) or (0.3, 0.5) representing effective and relatively ineffective interaction term to control the simulations for interaction term effect size. This simulation study is controlled for one sample size n=500 and ICC value, ICC=0.33. Dichotomous dummy variables, d_1 and d_2 are again used to include a time-independent categorical variable with three levels as in Section 5.1.1.2. Fitted models are labeled by M1 and M2 and are as follows

$$\text{M1: } \mu_{ij} = b_{0i} + 2\arctan(\beta_1 x_{ij} + \beta_2 d_{1i} + \beta_3 d_{2i} + \beta_4 x_{ij} d_{1i} + \beta_5 x_{ij} d_{2i}),$$

$$\text{M2: } \mu_{ij} = b_{0i} + 2\arctan(\beta_1 x_{ij} + \beta_2 d_{1i} + \beta_3 d_{2i})$$

The results of this simulation study are presented in Table 5.34. According to the results of Table 5.34, when the size of interaction effect is small, $PAPE_2$ outperforms the other criteria. Performances of $PAPE_1$ and CPD_2 are the same for ineffective interaction effects whereas CPD_1 performs worse. When effective interaction effects are concerned, CPD_1 and CPD_2 outperform $PAPE_1$ and $PAPE_2$. However, there is a little bit difference between their the frequency of selecting true model for low ICC (0.33).

Table 5.34: Frequency of selecting the true model. ICC=0.33, n=500.

(β_4, β_5)	(0.3,0.5)	(2.0,2.5)
CPD_1	74	96
$PAPE_1$	76	95
CPD_2	76	96
$PAPE_2$	79	95

5.3.2 Different Distributions

In this section, the results of Monte Carlo simulation study are presented to compare the performances of criteria, when data sets are generated from two different distributions. The logic behind data generation process is the same as in Section 5.1.2. ASC is used to study the performances of all criteria. This simulation study is controlled for two different sample sizes, n=500, 1000 and ICC values, ICC=0.33, 0.50. True models denoted by TM1 and TM2 are given in the hierarchical representation as follows

TM1: $\theta_{ij} \sim vM(\mu_{ij}, \kappa)$

$$\mu_{ij} = b_{0i} + 2\arctan(\beta_1 x_{ij}),$$

TM2: $\theta_{ij} \sim WC(\mu_{ij}, \rho)$

$$\mu_{ij} = (b_{0i} + \beta_1 x_{ij})\ [mod\, 2\pi]$$

where κ and ρ are set at 1.10 and 0.45 respectively so that the two distributions have the same circular dispersion, which is 2. Fitted model is TM1.

Table 5.35: Average Stability Coefficients over 100 Monte Carlo replications.

ICC	0.33		0.50	
n	500	1000	500	1000
$ASC(CPD_1)$	1.115	1.120	1.117	1.118
$ASC(PAPE_1)$	1.200	1.204	1.159	1.164
$ASC(CPD_2)$	1.149	1.157	1.152	1.154
$ASC(PAPE_2)$	1.264	1.269	1.208	1.214

Table 5.35 presents ASC of each criterion under different sample sizes and ICC values. According to results of Table 5.35, $PAPE_2$ can select more strongly the true model since ASC of this method is higher than those of the others irrespective of sample size and ICC. In terms of ASC, performances of CPD_1 and CPD_2 are close to each other for all sample sizes and ICC values. Overall, $PAPE_2$ seems to perform more satisfactorly when it is appropriate to use these methods. It is more decisive in its selection compared to others.

CHAPTER 6

APPLICATION

In this chapter, we illustrate the considered model and model selection on a real data set consisting of sandhopper orientations.

6.1 Data Description

Borgioli et al. (1999) and D'Elia et al. (2001) presented a longitudinal study in order to determine factors affecting the escaping mechanism of sandhoppers. In the original study, there were 144 sandhoppers but we are only able to retrieve a subset of the data. This subset contains 65 sandhoppers. Escape directions of each animal with respect to North were recorded 5 times (every 10 minutes), i.e equally-spaced repeated measurement study, with some covariates such as wind direction, the length of the left and right ocular diameters, and sun azimuth. Wind direction and sun azimuth variables are circular continuous variables. In our considered model, covariates should be linear, therefore, for modeling purposes the circular continuous variable wind direction is transformed into a categorical variable with four categories: wind from land [337°, 66°], wind from longshore-east [67°, 156°], wind from sea [157°, 246°], and wind from longshore-west [247°, 336°], with wind from the land taken as the reference category. Eye symmetry index (Eye) is constructed from ocular diameters, as

$$Eye = \log(\frac{max.diam.right \ \times \ min.diam.right}{max.diam.left \ \times \ min.diam.left}). \tag{6.1}$$

When Eye = 0, both eyes are equally wide for animal, Eye > 0 (or < 0) when right eye (left eye) is wider, respectively. Finally, the circular continuous variable sun azimuth is transformed into a categorical one with two categories: Morning [124°, 149°] and Afternoon [240°, 269°]. It is expected that animals will escape towards the sea and theoretical escape direction (TED) (the sea direction) is at $201°$.

6.2 Exploratory Analysis

First we summarize the main characteristics of data set ignoring for a moment the longitudinal structure of the data. Figure 6.1 represents the circular frequency distribution for each release separately. As seen from Figure 6.1 each marginal distribution of the escape directions for each release seems as a symmetric and unimodal distribution.

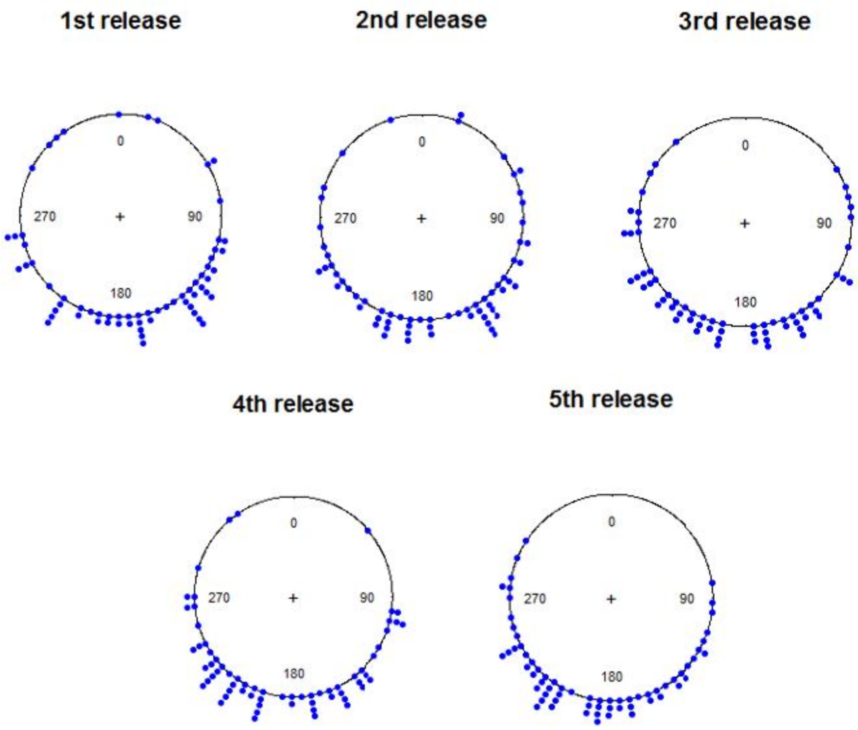

Figure 6.1: Circular frequency distributions of escape directions in 5 releases.

Circular summary statistics including the mean direction, mean resultant length, circular variance, circular symmetry coefficient and p-values of large-sample test for circular symmetry coefficient for the escape direction at each release are presented in Table 6.1. Results of Table 6.1 show that the mean direction ($\bar{\theta}$) of each release is close to $201°$ (TED) and there is a gradual approach to the TED (excluding 4^{th} release). There is an increase in mean resultant lengths ($\bar{R}$). Circular variance (V) decreases in subsequent releases. In other words, as the lag increases the dispersion around the mean direction decreases. Circular symmetry coefficient (s) verifies that each marginal distribution is a symmetric and unimodal distribution since circular symmetry coefficient is close to 0 for each release. A large-sample test for reflective symmetry (Pewsey, 2002) is performed to investigate whether the null hypothesis of circular reflective symmetry is rejected. Since all p-values (Table 6.1) are greater than 0.05, we can say that each marginal distribution is a symmetric distribution.

Table 6.1: Circular summary statistics for each marginal distribution (at each release).

Release	$\bar{\theta}$	$\bar{R}$	V	s	p-value
1^{st}	167.088	0.523	0.477	-0.244	0.283
2^{nd}	171.401	0.528	0.472	0.026	0.912
3^{rd}	193.242	0.576	0.424	0.098	0.733
4^{th}	190.887	0.627	0.373	0.170	0.582
5^{th}	194.585	0.667	0.333	0.204	0.543

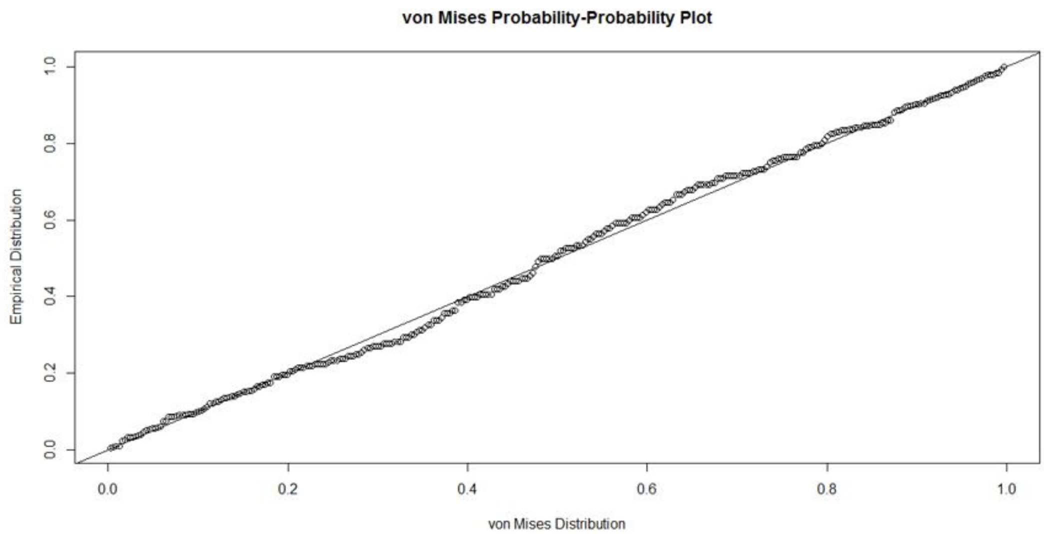

Figure 6.2: Plot of the empirical distribution of data set against von Mises distribution.

A Watson's goodness of fit test (Stephens, 1970; Jammalamadaka and SenGupta, 2001) is performed to investigate whether vM is a good fit for the data that consist of all the 5 jumps together. Test statistic is calculated as 0.0883 and p-value is equal to 0.092 (at significance level 0.05). Hence we can say that a vM distribution is a good fit for the data. Moreover, Figure 6.2 confirms that this data set seems to follow a vM distribution.

Since repeated measurements are taken from same animal, these five escape directions at each release are correlated. Thereby, the longitudinal structure of data set should be considered while analyzing it. First of all, a circular-circular correlation coefficient (Jammalamadaka and Sarma, 1988) is used to calculate correlation between successive measurements for same individual (autocorrelation coefficient). Table 6.2 represents circular autocorrelation coefficients for escape directions. It is clear from Table 6.2 that, there is a correlation between successive releases and as the lag increases, autocorrelation within same animal decreases (except 4^{th} release).

Table 6.2: Autocorrelation coefficient for escape directions.

Release	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
1^{st}	1	0.76	0.58	0.61	0.56
2^{nd}		1	0.70	0.67	0.68
3^{rd}			1	0.77	0.69
4^{th}				1	0.87
5^{th}					1

Longitudinal plot (spaghetti plot) of escape directions (in degree) at all releases for each animal is presented in Figure 6.3. Variability within any animal over time and variability between animals at any release can be explored by using this spaghetti plot. In this plot, the variability in each line represents variability within each sandhopper. Variability between animals represented by the space between lines.

6.3 Modeling

Borgioli et al. (1999), D'Elia (2001), Antonio and Pena (2014), Lagona (2016) and Maruotti (2016) previously worked on this data set. D'Elia (2001) used a variance component model under a vM distribution assumption. Antonio and Pena (2014)

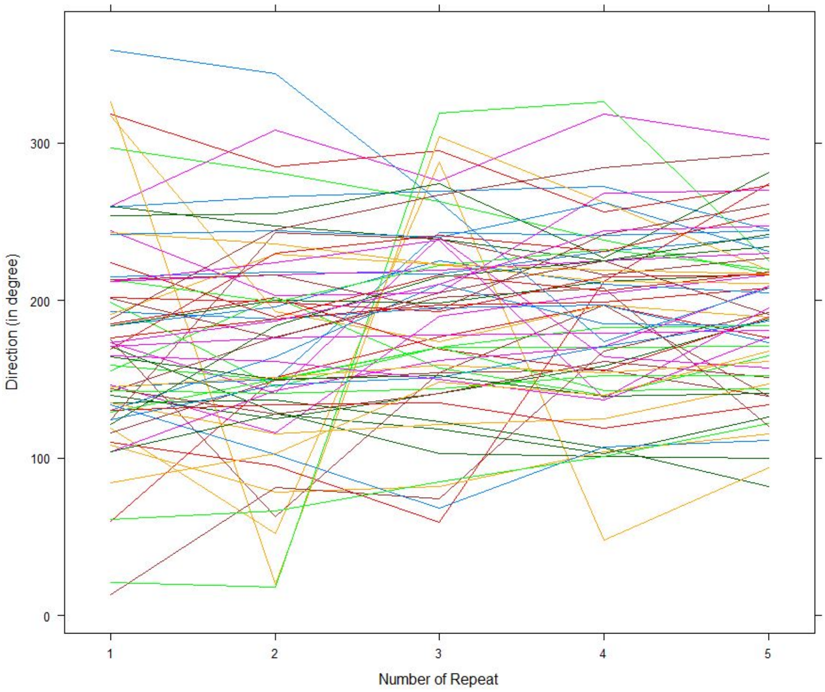

Figure 6.3: Longitudinal plot of escape directions for each sandhopper.

employed a projected circular longitudinal model under the projected bivariate normal distribution assumption (in Bayesian setting). Lagona (2016) exploit a mixed-effects model under the multivariate vM distribution assumption. Maruotti (2016) used a mixed-effects model under the projected bivariate normal distribution assumption.

In order to analyze this data set, four types of nested models are considered and represented in Table 6.3:

Table 6.3: Nested Models

Covariates	Mean Models
Sun	$\mu_{ij} = \beta_0 + 2\arctan(b_{0i} + \beta_1 Sun + \beta_2 Time)$
Sun + Eye	$\mu_{ij} = \beta_0 + 2\arctan(b_{0i} + \beta_1 Sun + \beta_2 Eye + \beta_3 Time)$
Sun + Wind	$\mu_{ij} = \beta_0 + 2\arctan(b_{0i} + \beta_1 Sun + \beta_2 Lse + \beta_3 Sea + \beta_4 Lsw + \beta_5 Time)$
Sun + Wind + Eye	$\mu_{ij} = \beta_0 + 2\arctan(b_{0i} + \beta_1 Sun + \beta_2 Lse + \beta_3 Sea + \beta_4 Lsw + \beta_5 Eye + \beta_6 Time)$

Here, β_0 is an offset parameter within $(-\pi, \pi)$, b_{0i} is subject specific random intercept for subject i which follows a normal distribution with mean zero and variance $\sigma^2_{b_0}$. The values of all criteria are calculated for each model and presented in Table 6.4. According to results of Table 6.4, CPD_1 and CPD_2 suggest that "Sun+Eye" model should be preferred, on the other hand $PAPE_1$ suggests that "Sun+Wind+Eye" model should be preferred. $PAPE_2$ select "Sun+Wind" model as the best fitting model to sandhopper data set. Based on the results of Table 6.4, the suggestion of CPD_1 and CPD_2 are consistent with the previous analysis of this data set in the literature. As seen in the simulation studies, since some factors can affect the performances of criteria, especially the performance of $PAPE_1$ and $PAPE_2$, we select "Sun+Eye" model (the selection of CPD_1 and CPD_2) as the best fitting model for sandhopper data set.

Table 6.4: Model comparison

	Models			
Tools	S	S+E	S+W	S+W+E
CPD_1	206.900	**206.000**	206.800	206.100
$PAPE_1$	390.004	390.237	389.216	**389.153**
CPD_2	90.520	**89.870**	90.490	89.910
$PAPE_2$	231.055	231.110	**230.184**	230.329

For "Sun+Eye" model, parameter estimations, posterior standard deviations, MC errors and 95% credible intervals are presented in Table 6.5. Parameter estimations are again consistent with previous analysis of this data set. MC errors of all parameters are less than 5% of their posterior standard deviations. All parameters of the model are significant. It is clear that the random intercept should be included in the model since the random intercept variance is significantly different from zero (0.270, 0.755). This means that there is variability between animals and random intercept may account for this variability. The MCMC convergence diagnostics for the selected model ("Sun+Eye") are given in Appendix B.

Table 6.5: Parameter Estimates

	Par. Est.	Std.	MC Error	95% Credible Int.
β_0	2.871	0.083	0.002	(2.701, 3.031)
Sun	0.239	0.096	0.003	(0.052, 0.428)
Eye	-1.101	0.101	0.002	(-1.300, -0.904)
Time	0.054	0.013	0.0003	(0.028, 0.080)
κ	3.655	0.278	0.005	(3.125, 4.213)
$\sigma^2_{b_0}$	0.463	0.124	0.002	(0.270, 0.755)

Interpretations of parameters are clear enough. For instance, interpretation of Sun coefficient is as follows. For simplicity, let Eye covariate be 0 and Time covariate be 1. Then,

for Sun=0, $\mu_{ij} = 2.871 + 2\arctan(0.239 \times 0 - 1.101 \times 0 + 0.054 \times 1) \approx 170.168°$

for Sun=1, $\mu_{ij} = 2.871 + 2\arctan(0.239 \times 1 - 1.101 \times 0 + 0.054 \times 1) \approx 197.097°$.

This means that when Sun=1, there will be an increase as $2\arctan(0.239) \approx 26.929°$ in mean direction of sandhopper. Other parameters can be interpreted in a similar way.

CHAPTER 7

CONCLUSION

In this thesis, modeling and model selection for repeated and circular observations (i.e. longitudinal circular data) which arise in many different areas such as biology, meteorology, medicine, biology, and geology have been investigated. Taking into account this data structure, a random intercept Bayesian model for longitudinal circular responses has been considered. The main benefits of this model is twofold: *i.* it overcomes the high dimensionality problem with multivariate distributions when within sample size is large, *ii.* it avoids the projected normal distribution based modeling that doubles the dimension of the latent random intercept space. We considered Bayesian approach which is particularly useful to deal with random intercept models. The major benefits of Bayesian in random intercept model is twofold: *i.* it treats the parameters and the latent random intercept within the same framework, *ii.* estimation of random intercept, test of random intercept variances are readily available in the posterior output. Considered Bayesian longitudinal circular random intercept model is working under the assumption of vM distribution but the Bayesian methodology can be adjusted for the other possible circular distributions. A detailed analysis of considered model has been provided in Bayesian framework. Full conditional distributions have been derived for each parameter and Gibbs sampling algorithm has been used to draw samples from the marginal posterior distributions. Bayesian estimators have been employed to efficiently estimate the model parameters. The performances of parameter estimators have been evaluated by using RB and MSE under correctly specified models. These measures also have been used to assess robustness of the Bayesian estimators to model misspecification. In terms of Bayesian parameter estimators, the main results of this thesis are listed below:

- Inferences for regression coefficients and concentration parameter are not robust to model misspecification irrespective of sample size, ICC value, and size of quadratic and interaction terms.

- Robustness of inference of random intercept parameters depends on many factors including quadratic and interaction effect sizes, sample size, and ICC value.

- Inferences of all parameters are not robust to distribution misspecification irrespective of sample size and ICC value.

- According to overall results, RB and MSE of Bayesian estimators are at acceptable level for all simulation scenarios.

On the other side, new methods based on minimizing a predictive loss have been considered to assess, compare and select the best fitting model given a set of candidate models. Their performances have been compared for a current model selection method (CPD_1). The frequency of selecting the true model has been used to evaluate the performances of methods under model misspecification. A new quantity called Average Stability Coefficient (ASC) has been introduced to assess the stability of methods on a particular selection under distribution misspecification. In terms of model assessment, comparison and selection, the results of this thesis are as follows:

- The performance of criteria depends on many factors including sample size, ICC value, the size of quadratic and interaction terms.

- Whether an intercept parameter is defined on the unit circle instead of a linear intercept parameter defined in the link function can also affect the performances of criteria.

- The performances of CPD_1 and CPD_2 are comparable irrespective of sample size, ICC value and the size of quadratic and interaction terms. The performances of $PAPE_1$ and $PAPE_2$ are likewise comparable for all simulation scenarios.

- According to general results of the simulation study, CPD_1 and CPD_2 have better performance than the others.

When uncertainty in the estimator is large and the function to be plugged-in (in our case APE) is nonlinear, errors from using plug-in (in our case $PAPE_1$ and $PAPE_2$) may be large (Rossi et. al, 2005). This means that the size of uncertainty in the estimator affects the performance of plug-in estimators. (For instance, on another context, in terms of estimating MSE, Maithy and Sherman (2008) showed that plug-in estimators perform inferior compared to bootstrap method in adaptive linear regression). For the correct evaluation of performances of these estimators ($PAPE_1$ and $PAPE_2$), uncertainty in prediction of $\hat{\theta}^{pred}$ and theoretical properties of these tools should be further investigated. For this reason, CPD_1 and CPD_2 should be preferred to determine the best fitting model when model selection for longitudinal circular response is concerned.

REFERENCES

[1] Abramowitz, M., and Stegun, I.A. (1965). *Handbook of Mathematical Functions*. Dover Publications, New York.

[2] Agostinelli, C., and Lund, U. (2011). Circular: Circular Statistics, https://cran.r-project.org/web/packages/circular/index.html

[3] Akaike, H. (1973). Information theory and the maximum likelihood principle. In *International Symposium on Information Theory*, Eds. Petrov, B.N., and Csaki, F. 267–281, Budapest: Akademiai Kiado.

[4] Artes, R., and Jorgensen, B. (2000). Longitudinal data estimating equations for dispersion models. *Scandinavian Journal of Statistics* , 27, 321–334.

[5] Artes, R., Paula, G.A., Ranvaud, R. (2000). Analysis of circular longitudinal data based on generalized estimating equations. *Australian and New Zealand Journal of Statistics*, 42, 347–358.

[6] Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19, 3-11.

[7] Bartko, J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin*, 83, 762-765.

[8] Carnicero, J.A., and Wiper, M.P. (2008). A semi-parametric model for circular data based on mixtures of beta distributions. *Statistic and Econometric Series*, 08-05.

[9] Celeux G., Forbes F., Robert C.P., and Titterington D.M. (2006). Deviance Information Criteria for Missing Data Models. *Bayesian Analysis*, 1, 651–674

[10] Çamlı, O., Kalaylıoğlu, Z. (2016). Model Selection Criteria for Longitudinal Circular Data Analysis. International Conference on Information Complexity and Statistical Modeling in High Dimensions with Applications.

[11] D'Elia, A. (2001). A statistical model for orientation mechanism. *Statistical Methods and Applications*, 10, 157–174.

[12] Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S.L. (1994). *Analysis of Longitudinal Data*. Oxford University Press, Oxford.

[13] Diggle, P.J., Heagerty P., Liang K.Y., and Zeger, S.L. (2002). *Analysis of Longitudinal Data, second edition*. Oxford University Press, Oxford.

[14] Downs, T.D., and Mardia, K.V. (2002). Circular regression. *Biometrika*, 89, 683–697.

[15] Feldt, L.S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, 30, 357-370.

[16] Field, A.P. (2005). Intraclass correlation. In *Enclyopedia of Statistics in Behavioral Sciences*, Eds. Everitt, B.S., and Howell, D.C. 30–46, Chichester, John Wiley and Sons, Ltd.

[17] Fisher, N.I., and Lee A.J. (1983). A Correlation Coefficient for Circular Data. *Biometrika*, 70, 327-332

[18] Fisher, N.I., and Lee A.J. (1992). Regression models for angular response. *Biometrics*, 48, 665–677.

[19] Fisher, N.I. (1993). *Statistical Analysis Circular Data*. Cambridge University Press, Cambridge.

[20] Fitzmaurice, G., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*, Wiley-Interscience, New Jersey.

[21] Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008). *Longitudinal Data Analysis*. Chapman Hall, New York.

[22] Geisser, S., and Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 153–160.

[23] Gelfand, A.E., and Ghosh, S.K. (1998). Model choice: A minimum posterior predictive loss function. *Biometrika*, 85, 1–11.

[24] Gelman, A., Hill, J. (2007). *Data Analysis Using Regression and Multilevel Hierarchical Models*. Oxford University Press, New York.

[25] George, B. J. and Ghosh, K. (2006).A Semiparametric Bayesian Model for Circular-Linear Regression. *Communications in Statistics - Simulation and Computation,* 35:4, 911-923.

[26] Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman Hall, London.

[27] Griffin, D., and Gonzalez, R. (1995). Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*, 118, 430-439.

[28] Haggard, E.A. (1958). *Intraclass correlation and the analysis oj variance*. Dryden Press, New York.

[29] Hall, D.B., Shen J. (2015). Marginal projected multivariate linear models for clustered angular data. *Australian and New Zealand Journal of Statistics*, 57(2), 241–257.

[30] Hayes, W.L. (1973). *Statistics for the social sciences*. New York: Holt, Rinehart and Winston.

[31] Hedeker, D., and Gibbons, R.D. (2006). *Longitudinal data analysis*. Wiley-Interscience, New Jersey.

[32] Ilk, O. (2008). Multivariate Longitudinal Data Analysis: Models for Binary Response and Exploratory Tools for Binary and Continuous Response. Verlag Dr. Muller (VDM), Verlag.

[33] Jammalamadaka, S., and Sarma, Y. (1988). *A correlation coefficient for angular variables.* Statistical Theory and Data Analysis 2. North Holland: New York.

[34] Jammalamadaka R.A., and SenGupta, A. (2001). *Topics in circular statistics*. World Scientific, Singapore.

[35] Jorgensen, B. (1997). Proper dispersion models (with discussion). *The Brazilian Journal of Probability and Statistics*. 11, 8140.

[36] Jorgensen, B. (1997). *The theory of dispersion models*. Chapman Hall, London.

[37] Kent, J. T. and Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics*. 15, 247-54.

[38] Lagona, F. (2016). Regression analysis of correlated circular data based on the multivariate von Mises distribution. *Environmental and Ecological Statistics*, 23(1), 89-113.

[39] Liang,K-Y.,and Zeger, S.L. (1986). Longitudinal analysis using generalized linear models. *Biometrika*, 73, 13–22.

[40] Maity, A., Sherman, M. (2008). On adaptive linear regression. *Journal of Applied Statistics,* 35:12, 1409-1422.

[41] Mardia, K.V., and Jupp, P.E. (2000). *Directional statistics*. New York: John Wiley and Sons.

[42] Martinez, W.L., and Martinez, A.R.(2002). *Computational Statistics Handbook with MATLAB*, CRC Press, New York.

[43] Maruotti, A., Punzo A., Mastrantonio, G., and Lagona, F. (2016). A time-dependent extension of the projected normal regression model for longitudinal circular data based on a hidden Markov heterogeneity structure. *Stochastic Environmental Research and Risk Assessment*, 30, 1725–1740.

[44] Maruotti, A. (2016). Analyzing longitudinal circular data by projected normal models: a semi-parametric approach based on finite mixture models. *Stochastic Environmental Research and Risk Assessment*, 23, 257-277.

[45] McGraw, K.O., and Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1, 30-46.

[46] McMillan, G.P., Hanson, T.E., Saunders, G., Gallun, F.J. (2013). A two-component circular regression model for repeated measures auditory localization data. *Series C Applied Statistics*, 62(4), 515-534.

[47] McMillan, G.P., Saunders, G., and Hanson, T.E. (2011). A statistical model of horizontal auditory localization performance data. *The Journal of the Acoustical Society of America,*, 129, EL229–EL235.

[48] Nunez-Antonio, G., and Gutierrez-Pena E. (2014). A Bayesian model for longitudinal circular data based on the projected normal distribution. *Computational Statistics and Data Analysis*, 71, 506–519.

[49] Pewsey, A. (2002). Testing circular symmetry. *Canadian Journal of Statistics*, 30, 591–600.

[50] Presnell. B, Morrison, S.P., and Littell, R.C. (1998). Projected multivariate linear model for directional data. *Journal of the American Statistical Association*, 93, 1068–1077.

[51] R Development Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[52] Ravindran, P.K., and Ghosh, S.K. (2011). Bayesian analysis of circular data using wrapped distributions. *Journal of Statistical Theory and Practice*, 5, 547-561.

[53] Rueda, C., Fernandez, M.A., Barragan, S., Mardia, K.V., and Peddada, S.D. (2016). Circular piecewise regression with applications to cell-cycle data. *Biometrics*, 72, 1266-1274.

[54] Scapini F. (1997). Variation in scototaxis and orientation adaptation of Talitrus saltator populations subjected to different ecological constraints. *Estuarine, Coastal and Shelf Science*, 44, 139–146.

[55] Small, C.G. (1996). *The Statistical Theory of Shape*, Springer, Verlag-New York.

[56] Snedecor, G.W., and Cochran, W.G. (1967). *Statistical methods (6th ed.)*. Iowa State University Press, Ames.

[57] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A. (2002). Bayesian measures of model complexity and fit, (with discussion and rejoinder). *Journal of the Royal Statistical Society*, Series B, 64, 583-639.

[58] Stephens, M. (1970). Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society*, B32, 115-122.

[59] Von Mises, R. (1918). Über die “Ganzzahligkeit” der Atomgewicht und verwandte Fragen. Physikalische Zeitschrift. 19, 490–500.

[60] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika*, 61, 439–447.

[61] Winer, B.J. (1971). *Statistical principles in experimental Design (2nd ed.)*. McGraw-Hill, New York.

[62] Ye, J. (1998). On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, 93, 120–131.

[63] Yuce T, Kalafat E, and Koc A. (2015). Transperineal ultrasonography for labor management: accuracy and reliability. *Acta Obstet Gynecol Scand* , 94, 760–5.

[64] Zhang, J., Troendle, J., Mikolajczyk, R., Sundaram, R., Beaver, J., Fraser, W. (2010). The Natural History of the Normal First Stage of Labor. *Obstetrics Gynecology*, 115(4) 705–710.

APPENDIX A

DIRECTED GRAPHICAL MODEL (DAG) FOR LCRIM

Figure A.1 represents the directed graphical model (DAG) for LCRIM. In this graphical model, solid arrows denote a stochastic relationship, hollow arrows indicate logical relationship. Stochastic nodes (variables) and constants are represented as ellipses and rectangles respectively in the DAG. Loops (repeated parts) in the model are denoted by plates.

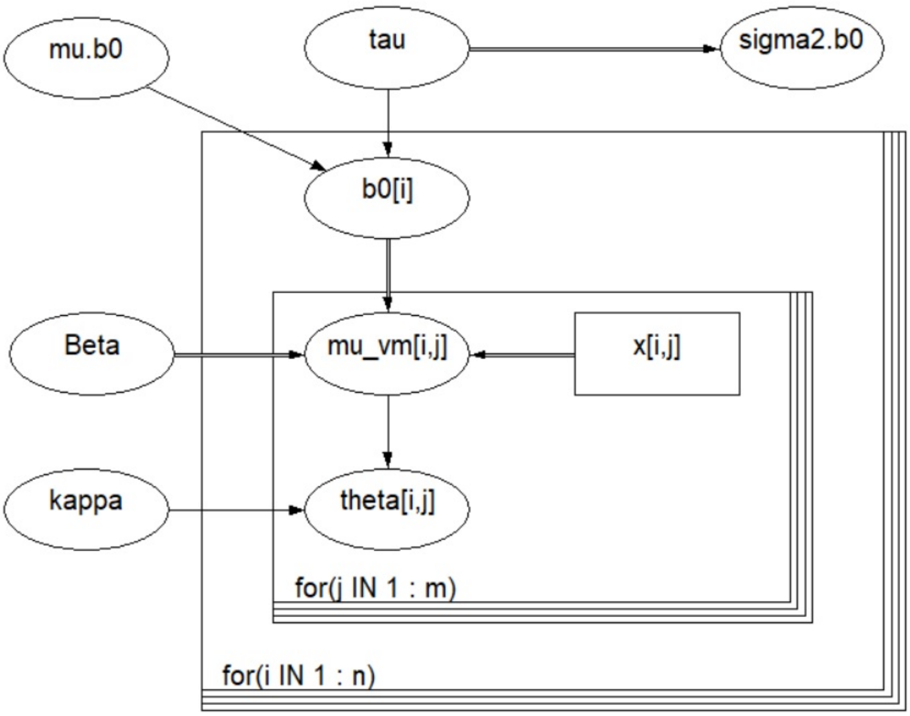

Figure A.1: DAG of LCRIM.

APPENDIX B

MCMC CONVERGENCE DIAGNOSTICS FOR "SUN+EYE" MODEL

In order to assess convergence in Gibbs Sampling, we used the following practical guidelines by starting two different chains with different initial values;

- History plots of the sample values versus iteration,
- Brooks-Gelman-Rubin (BGR) statistics (Brooks and Gelman, 1998).

We checked convergence for each parameter of the model. Figures B.2 and B.1 show the history plots of all the parameters for "Sun+Eye" model. Figure B.3 represents the BGR statistics for the model of interest. In order to investigate the slow convergence, we also provided the autocorrelation plots indicating the serial correlation in the chains. The autocorrelation plots are presented in Figure B.4.

According to Figures B.2 and B.1, we can be reasonably confident that convergence has been achieved since all the chains appear to be overlapping one another. Figure B.3 shows that convergence has been achieved after about 10000 iteration as the BGR statistics for all parameters tend to 1 and converge to stability. We determined the burn-in period of the Markov chain by using these figures for the LCRIM.

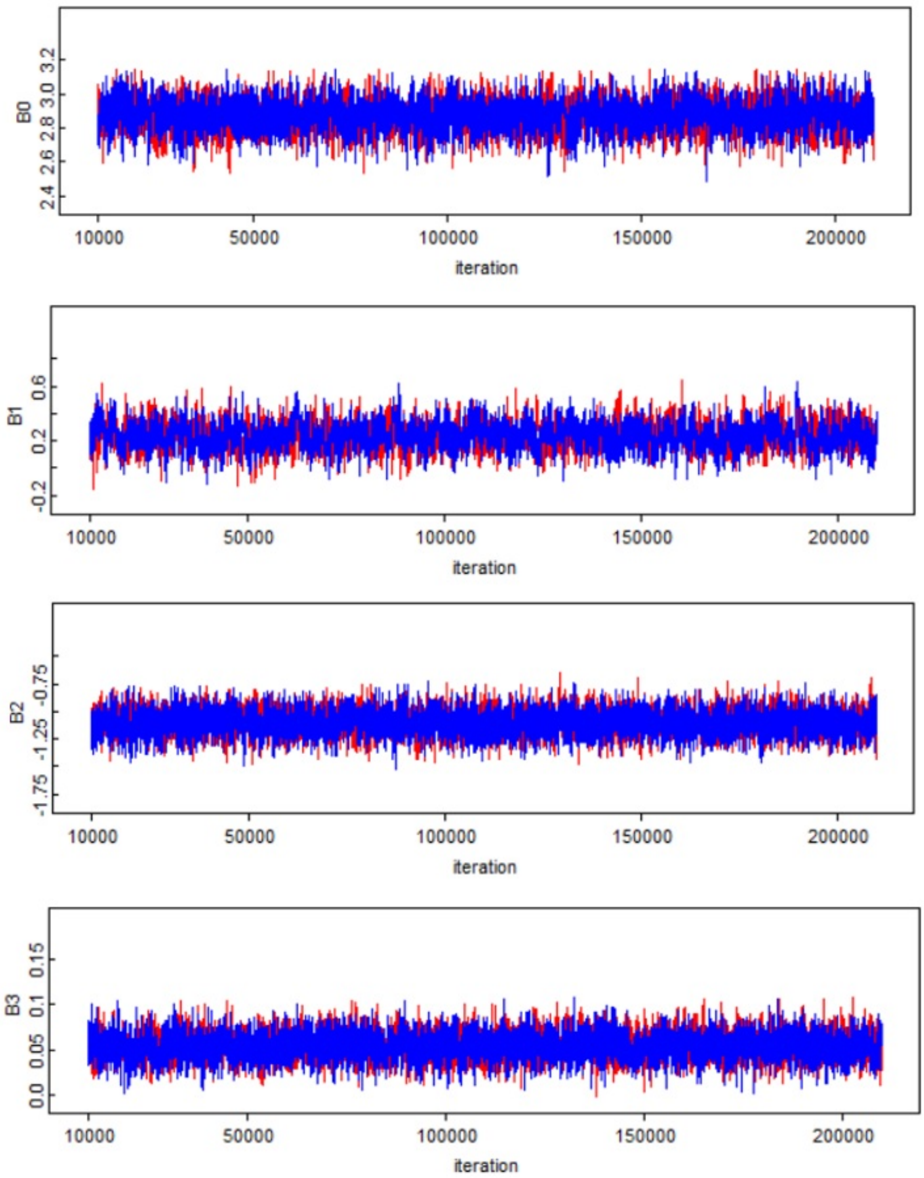

Figure B.1: The History Plots for the regression coefficients.

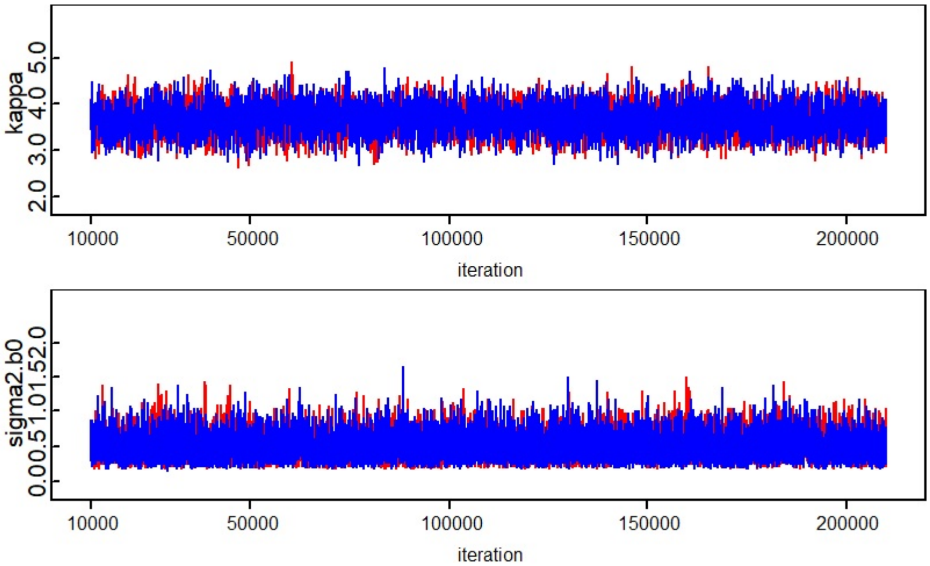

Figure B.2: The History Plots for κ and $\sigma^2_{b_0}$.

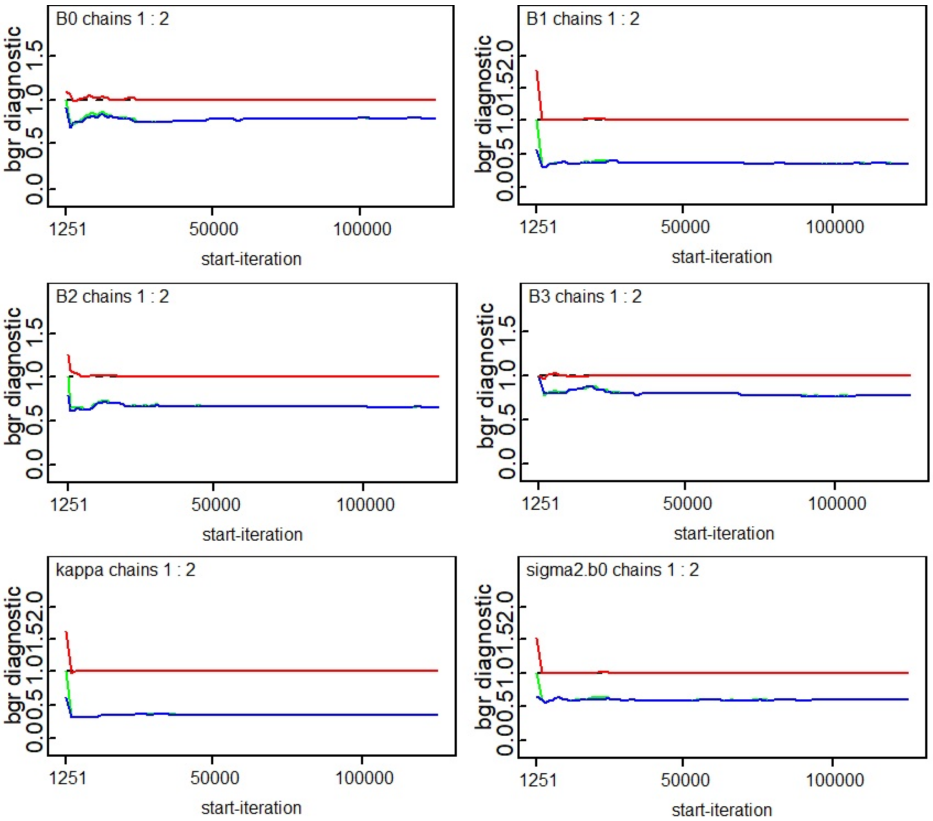

Figure B.3: The Brooks-Gelman-Rubin's Convergence Diagnostics for "Sun+ Eye" Model.

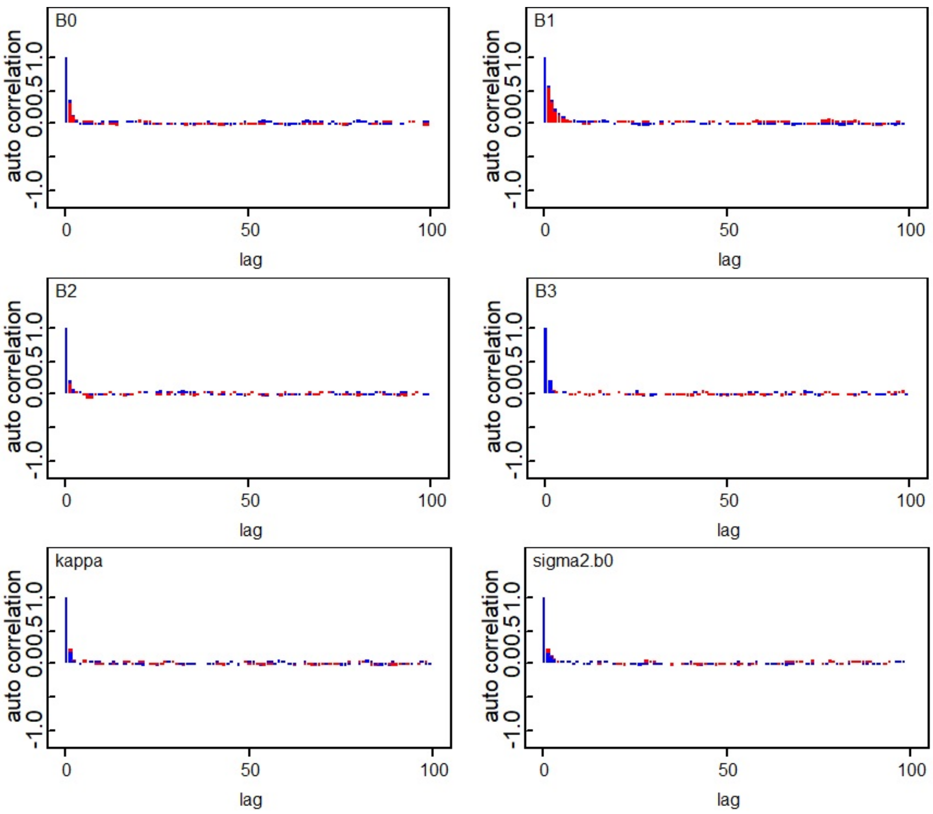

Figure B.4: The Autocorrelation Plots for "Sun+ Eye" Model.