BAYESIAN VARIABLE SELECTION IN CIRCULAR REGRESSION MODELS USING LASSO

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 DOCTOR OF PHILOSOPHY
IN
STATISTICS

SEPTEMBER 2023

Approval of the thesis:

BAYESIAN VARIABLE SELECTION IN CIRCULAR REGRESSION MODELS USING LASSO

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

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Özlem İlk Dağ
Head of Department, **Statistics** _____

Prof. Dr. Zeynep Kalaylıoğlu
Supervisor, **Statistics, METU** _____

Prof. Dr. Ashis SenGupta
Co-supervisor, **Indian Statistical Institute, Kolkata, India** _____

Examining Committee Members:

Prof. Dr. Çağdaş Hakan Aladağ
Statistics, Hacettepe University _____

Prof. Dr. Zeynep Kalaylıoğlu
Statistics, METU _____

Prof. Dr. Özlem İlk Dağ
Statistics, METU _____

Prof. Dr. Barış Sürücü
Statistics, METU _____

Assoc. Prof. Dr. Muhammet Burak Kılıç
Business Administration, MAKU _____

Date:11.09.2023

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, Surname: Onur Çamlı

Signature :

ABSTRACT

BAYESIAN VARIABLE SELECTION IN CIRCULAR REGRESSION MODELS USING LASSO

Çamlı, Onur
Ph.D., Department of Statistics
Supervisor: Prof. Dr. Zeynep Kalaylıoğlu
Co-Supervisor: Prof. Dr. Ashis SenGupta

September 2023, 121 pages

Applications of circular regression models are ubiquitous in many disciplines, particularly in meteorology, biology and geology. In these models, variable selection problem continues to be a remarkable open question. In this thesis, inspired by the Bayesian lasso used in Euclidean space, we consider new shrinkage-based approaches for variable selection in circular regression models. Firstly, we adapt Bayesian lasso for linear-circular regression models and show that it is not able to produce robust inference as the coefficient estimates are sensitive to the choice of hyper-prior setting for the tuning parameter. To eradicate the problem, we propose a robustified Bayesian lasso that is based on an empirical Bayes (EB) type methodology to construct a hyper-prior for the tuning parameter. This construction is computationally feasible and can be used effectively in both Euclidean and circular regression settings. Simulation studies show that robustified Bayesian lasso leads to a more robust inference. Then, we compare the performance of our proposed method and the existing prior distributions in the literature using two different real data sets. Overall, the method offers an efficient Bayesian lasso for variable selection in linear-circular regression. Secondly,

we adapted Bayesian lasso and Bayesian adaptive lasso methods that can allow for individual shrinking coefficients for each coefficient in circular-(circular, linear) regression models. We also develeoped some alternatives of Bayesian adaptive lasso based on Dirichlet Process prior. In this context, we provide a comprehensive simulation study to show the performances and behavior of the adapted and proposed methodologies in terms of parameter estimation and variable selection. The results indicate that the different methods being compared have similar performance based on evaluation metrics for both parameter estimation and variable selection in circular-(circular, linear) regression models.

Keywords: Variable selection, Bayesian lasso, circular regression, dimension reduction

ÖZ

DAİRESEL REGRESYON MODELLERİNDE LASSOYA DAYALI BAYESÇİ DEĞİŞKEN SEÇİMİ

Çamlı, Onur
Doktora, İstatistik Bölümü
Tez Yöneticisi: Prof. Dr. Zeynep Kalaylıoğlu
Ortak Tez Yöneticisi: Prof. Dr. Ashis SenGupta

Eylül 2023 , 121 sayfa

Dairesel regresyon modellerinin uygulamaları, meteoroloji, biyoloji ve jeoloji gibi birçok disiplinde yaygındır. Bu modellerde, değişken seçimi problemi dikkate değer bir açık soru olmaya devam etmektedir. Bu tezde, Öklid uzayında kullanılan Bayesçi lasso'dan esinlenerek, dairesel regresyon modellerinde değişken seçimi için yeni küçülme tabanlı yaklaşımları ele aldık. İlk olarak, Bayesçi lasso'yu doğrusal-dairesel regresyon modelleri için uyarlıdık ve katsayı tahminlerinin ayar parametresi için hiper-önsel dağılımının seçimine duyarlı olması nedeniyle sağlam çıkarım üretemediğini gösterdik. Sorunu ortadan kaldırmak için, ayar parametresi için bir hiper-önsel oluşturmak üzere ampirik Bayes (EB) tipi bir metodolojiye dayanan sağlamlaştırılmış bir Bayesçi lasso önerdik. Bu yaklaşım hesaplama açısından uygundur ve hem Öklid hem de dairesel regresyon ortamlarında etkili bir şekilde kullanılabilmektedir. Simülasyon çalışmaları, sağlamlaştırılmış Bayesçi lasso'nun daha sağlam bir çıkarıma yol açtığını göstermektedir. İki farklı gerçek veri seti kullanarak önerdiğimiz yöntemin performansını literatürdeki önsel dağılımların performansları ile kar-

şılaştırdık. Genel olarak, yöntem doğrusal-dairesel regresyonda değişken seçimi için etkili bir Bayesçi lasso sunmaktadır. İkinci olarak, dairesel-(dairesel, doğrusal) regresyon modellerinde her katsayı için bireysel küçültme katsayılarına izin verebilen Bayesçi lasso ve Bayesçi uyarlanabilir lasso yöntemlerini uyarladık. Ayrıca Dirichlet Süreci önceliğine dayalı Bayesçi uyarlanabilir lasso'nun bazı alternatiflerini geliştirdik. Uyarlanan ve önerilen metodolojilerin parametre tahmini ve değişken seçimi açısından performanslarını ve davranışlarını göstermek için kapsamlı bir simülasyon çalışması sunduk. Sonuçlar, karşılaştırılan yöntemlerin dairesel-(dairesel, doğrusal) regresyon modellerinde hem parametre tahmini hem de değişken seçimi için değerlendirme metriklerine dayalı olarak benzer performansa sahip olduğunu göstermektedir.

Anahtar Kelimeler: Değişken seçimi, Bayesci lasso, dairesel regresyon, boyut azaltma

To Royal Horses

ACKNOWLEDGMENTS

I am very grateful to my thesis advisor Prof. Dr. Zeynep 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 also like to thank my co-supervisor, Prof. Dr. Ashis SenGupta for his guidance and contribution to the initiation and progress of this research.

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

ACI	Average Correct Inclusion
AFE	Average False Exclusion
AIC	Akaike Information Criterion
APE	Average Prediction Error
AQI	Air Quality Index
BIC	Bayesian Information Criterion
BGR	Brooks-Gelman-Rubin
BaLasso	Bayesian Adaptive Lasso
CIs	Credible Intervals
CIR	Correct Inclusion Ratio
CMSE	Circular Mean-Square Error
CPD	Circular Predictive Discrepancy
$DLasso_1$	Dirichlet Lasso 1
$DLasso_2$	Dirichlet Lasso 2
DP	Dirichlet Process
EB	Empirical Bayes
EB-GS	Empirical Bayes Gibbs Sampling
EM	Expectation Maximization
GAIC	Generalized Information Akaike Criterion
GS	Gibbs Sampling
MSE	Mean Squared Error
MCMC	Markov Chain Monte Carlo
Lasso	Least absolute shrinkage and selection operator
LPML	Log Pseudo-Marginal Likelihood

OLS	Ordinary Least Squares
RB	Relative Bias
vM	Von Mises

CHAPTER 1

INTRODUCTION

Circular data, which also include periodic data and angular data, refer to the data that is measured on a circular scale or wrapped around a circle. This type of data is characterized by having values that repeat after a certain interval or period. A common example of circular data is angles, which range from 0 to 360 degrees or 0 to 2π radians, and repeating once the maximum value is reached. Circular data arise in various fields, including astronomy (e.g. celestial angles), meteorology (e.g. wind directions), biology (e.g. animal migration directions or directions of ants leaving their hill), time (e.g. time of the day when a patient's blood pressure reaches its peak). Various other examples can be found in Fisher (1993), Mardia and Jupp (1999) and Jammalamadaka and Sengupta (2001).

It is essential to handle circular data differently than traditional linear data because of its cyclical nature. One critical consideration when dealing with circular data is how to measure differences or distances between two circular values. Also, the usual arithmetic mean cannot be used for averaging angles because, for example, averaging 0 degrees and 359 degrees would give 179.5 degrees, which is not meaningful as it indicates opposite direction. Instead, specialized statistical methods like circular mean (also known as the mean resultant) and circular standard deviation are used to calculate summary statistics for circular data. In summary, circular data require special statistical techniques and visualizations to properly analyze and interpret the data.

An important concept in circular data analysis is circular regression models. Circular regression models are needed in many different fields such as meteorology, biology, geology, medicine, and psychology. There are a variety of real-life problems in

which circular regression models can be used to investigate the relationship between variables and to predict the dependent variable based on the covariates. A circular regression model refers to a regression model involving circular variables either as response or covariates. 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 linear-circular regression model that can be used to investigate the relationship between linear response and circular explanatory variables. For instance, Johnson and Wehrly (1978) considered a linear-circular regression model to predict the air quality index as a function of temperature, wind direction and wind speed by using their data set.

The second one is circular-linear regression model that can be used to investigate the relationship between circular response variable and a set of linear explanatory variables. In a circular-linear regression model, a circular variable is regressed onto one or more linear covariates. For example, it may be of interest to predict the direction of ocean current as a function of distance from the shore. Another example would be to predict the time of peak in blood pressure as a function of the amount of medication, which is useful in a review of dosing specifications.

The last one is circular-circular regression model, in which the variation of circular response variable can be explained in terms of some circular covariates. For instance, in a marine biology study, a data set was collected on the spawning times of a particular fish and the dependence of the spawning time on two explanatory variables one of which is time of low tide and the other one is amplitude of low tide were investigated (Lund, 1999). Such examples are found in Hrushesky (1985), Jones and Silverman (1989), Rivest (1997), Lowrey et al. (2000), Shearman et al. (2000), Mardia et al. (2008), Fernández-Durán and Gregorio-Domínguez (2014), Kim and SenGupta (2015), and Rueda et al. (2016). Circular regression models have been studied in Johnson and Wehrly (1978), Rivest (1997), Fisher (1992), Downs and Mardia(2002), SenGupta and Ugwuowo (2006), Kim (2009), SenGupta et al. (2013) and Kim and SenGupta (2016).

When regression models are of concern, an important problem is variable selection. Often, in the initial stage of a study, there are a large number of independent variables

which are potentially important for the model. To enhance predictability and give a parsimonious model, it is common practice to include only the significant covariates in the model. The problem of variable selection in regression models has received much attention recently, as variable selection is used for removing irrelevant, noisy or unreliable variables from the regression model to improve the model performance and predictions. The model complexity can be also reduced by removing these variables, in other words in some situations, the variable selection is used to obtain a model that is easier to understand. Also, variable selection can be used to reduce the risk of overfitting.

For these reasons, a number of variable selection techniques have been developed for regression models where the dependent variable is linear. However, there is no agreed upon "best" technique. These techniques differ in the algorithms used to identify subsets as well as in the criteria used to judge the goodness-of-fit of candidate models. The applicability of these methods depend on the data and the purpose of the study. Traditionally, variable selection in regression models for linear variables has been addressed using three broad classes of approaches. The first class is based on sequential statistical testing procedure (F, chi-square and t tests etc.). Briefly summarized, this approach reduces a general statistical model established to capture the characteristics of the data to a smaller model based on sequential testing. The second class of approaches is based on information theoretic ideas. Such ideas are generically known as information criteria or statistical criteria such as Adjusted R-squared, Mallows' Cp, MSE (Mean Squared Error), AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion). Generally, in these approashes, a score is provided by these criteria for every model established with different combination of the covariates. Some criteria also use a penalty term for model complexity. The model that optimises the criterion is chosen to be the best representation of the data. The last class for variable selection procedures is based on regularization (penalization) methods such as least absolute shrinkage and selection operator (lasso) (Tibshirani, 1996) received a fair bit of attention in the recent literature. Basically, the lasso aims to minimize the sum of squared errors as for least squares regression subject to the constraint that the L_1-norm of the regression coefficient vector should be below a predefined threshold. The degree of variable selection can be determined by the value of this

threshold. Furthermore, penalization is becoming increasingly popular in Bayesian framework. In Bayesian penalization, the prior distribution of regression coefficients are determined so that small effects are shrunken to zero while true large effects are maintained in the model. Bayesian penalization methods work sometimes better than classical penalization methods and they have some advantages such as automatic estimation of the penalty parameter, readily available uncertainty estimates and more flexibility in terms of penalties (Van Erp et al., 2019). Tibshirani (1996) notes that the lasso has a Bayesian interpretation. A Bayesian posterior mode can be used as a point estimation when the regression parameters have independent Laplace (i.e., double exponential (DE)) priors. The first explicit treatment of Bayesian lasso was provided by Park and Casella (2008). They used the scale mixture of normals representation of the double-exponential distribution to create a hierarchical formulation of the model.

There are many advantages in using the lasso method, first of all it can provide a very good prediction accuracy, because shrinking the estimated coefficients, we can often substantially reduce the variance at the cost of a negligible increase in bias, substantially improve the accuracy of prediction for future observations (James et al., 2013. This is especially useful number of observation is small and the number of covariates is large. Moreover, the lasso helps to increase the model interpretability by eliminating irrelevant variables that are not associated with the response variable, this way also overfitting is reduced. A wide variety of models can be fitted to large data sets by using the lasso thanks to newly developed computational algorithms. The lasso has statistical and computational gains. The availability of fast computation algorithms for estimating the lasso regressions in major programming and statistical languages, including Matlab, R, and SAS, suggest the usage of the method.

When variable selection in a circular regression is of concern, it is a major open research topic in statistics. The variable selection problems have not been well addressed for circular regression models. Our focus in this thesis is variable selection for circular regression model. In other words, our focus is on finding a parsimonious set of predictor variables which gives a good fit, predicts the dependent value well. Variable selection is intended to select the ”best” subset of predictors for circular regression model. In the following sections, we describe the open problem addressed in this thesis, previous work on the problem, and specific aims of the thesis.

1.1 Motivation and Problem Definition

In circular regression, just like in standard linear regression, the objective of variable selection is to identify the independent variables that have a significant impact on predicting the dependent variable, while considering the circular characteristics of the data. By selecting significant predictors, the model becomes more interpretable and easier to understand, and it can lead to better predictions. Variable selection in circular regression models is a rather fertile area for methodological and theoretical development.

Three significant works, Tibshirani (1996) (Lasso), Park and Casella (2008) (Bayesian Lasso), and Leng et al. (2014) (Bayesian Adaptive Lasso) have addressed variable selection in the context of regression models for linear variables. These methods can help in automatically selecting significant predictors by shrinking some regression coefficients towards zero. This leads to sparsity in the model, where only the significant predictors have non-zero coefficients, effectively achieving variable selection. In circular regression, these techniques need to be adapted to account for the circular nature of the circular variable.

Motivated these works, to address variable selection in circular regression models, this thesis proposes to explore a class of Bayesian penalization techniques such as Bayesian lasso and Bayesian adaptive lasso to reduce model complexity and the risk of over-fitting and improve the prediction ability of the model by giving a parsimonious model. In other words, we try to obtain a model that is easier to understand and get better predictions for circular regression models by developing a class of Bayesian penalization techniques based on Bayesian lasso and Bayesian adaptive lasso. It is important to note that, to the best of our knowledge, this is the first work that combines circular data and Bayesian lasso logic.

1.2 Objectives of the Research

Selection of an appropriate approximating model containing the significant variables is critical to statistical inference. In the literature, there are a lot of variable selection

methods for regression models in Euclidean space, but when modeling with circular variables is of concern, these usual variable selection methods 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 i.e. the circular sample space and the support of the data being a circle.

By applying a class of Bayesian penalization techniques such as Bayesian lasso and Bayesian adaptive lasso to circular regression models, this thesis aims to achieve the following benefits:

- Develop a new class of shrinkage-based variable selection techniques that take the account of circular properties of the data for circular regression models.
- Reduce model complexity. The use of regularization priors helps in controlling model complexity, leading to a more parsimonious model with fewer predictors, which is easier to interpret.
- Improve prediction ability. By selecting only the significant predictors, the model's predictive performance is likely to improve, as it focuses on the key factors driving the circular outcome.
- Evaluate and compare the performances of prior distributions for shrinkage parameters (tuning parameters) by considering their impact on the model's performance, such as predictive accuracy, model interpretability, and robustness (In linear-circular regression models).
- Illustrate the utilization of the developed shrinkage-based variable selection techniques by using some simulated and real circular data sets.

In summary, this thesis explores Bayesian penalization techniques to address variable selection in circular regression models, aiming to enhance model interpretability and predictive performance. These methods can be valuable for researchers and practitioners working with circular data and seeking more efficient and accurate modeling approaches. This study hopes to address and contribute to variable selection in circular regression models, a rather fertile area for methodological and theoretical development, while the demand increases with the advancing technology.

CHAPTER 2

LITERATURE REVIEW

In this chapter, an extensive literature review is provided on two main topics: circular regression models that is given in Section 2.1 and variable selection within these models which is given in Section 2.2.

2.1 Circular Regression Models

In this section, we consider circular regression models. Throughout the section, we use the generic notation of variables to denote that θ is a circular variable and X is linear variable. Circular regression models are statistical models designed to analyze circular data, which are measurements or observations that occur on a circular scale, such as angles, directions, or time of day. Unlike linear regression models that work with continuous data on a linear scale, circular regression models take into account the periodic nature of circular data.

The main challenge with circular data is that the values wrap around in a circular fashion, meaning that the highest value is immediately followed by the lowest value and vice versa. For example, in a 24-hour clock, 23:59 is immediately followed by 00:00. Therefore, traditional linear regression techniques would not be appropriate for circular data analysis.

In the literature, there are several regression models when the modeling with a circular variable is of concern. The variation of a circular response variable can be explained in a number of ways in terms of explanatory variables. 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. The focus of this study is to describe circular regression models for mean direction. These models assume that all the concentration parameters are equal to κ say, and the the mean directions μ_i's are related to the some covariates. There are several circular regression models used to deal with this kind of data, and some of the common ones are listed as follows;

2.1.1 Circular-Linear Regression Models

Circular-linear regression models are statistical models used when the relationship between a circular variable and a linear variable needs to be analyzed. In such cases, the response variable is measured in angles, often represented as points on the unit circle (radians or degrees), while the predictor variable is continuous or discrete and measured in linear units. The circular-linear regression model is designed to address the circular nature of the response variable and take into account its periodicity. This is essential because conventional linear regression models are not appropriate for circular data, as they do not account for the periodic wrapping around the circle.

These regression models also can be divided into some sub-classes depending on error distribution as follows; based on von Mises distribution (vM), the use of wrapped distributions, based on projected normal distribution and the non-parametric regresssion models. These approaches are the most known and applied approaches in circular data applications.

2.1.1.1 Based on von Mises Distribution

The circular-linear regression models were mostly built on the the vM distribution assumptions due to the established theoretical framework of this distribution and some key properties they inherit for statistical inference (Jones and Pewsey, 2005). This distribution is uni-modal and symmetric (characterized by its bell-shaped curve) and follows the following probability density function

$$f(\theta|\mu,\kappa) = [2\pi I_0(\kappa)]^{-1}\exp\{\kappa\cos(\theta-\mu)\}, \quad 0 \le \theta < 2\pi, \quad \kappa \ge 0 \tag{2.1}$$

where θ is circular random variable, μ is the mean direction and κ is the concentration parameter. Here, $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero and it is given by

$$I_0(\kappa) = (2\pi)^{-1}\int_0^{2\pi}\exp(\kappa\cos(\phi))d\phi. \tag{2.2}$$

Let $\theta_1, ..., \theta_n$ be a random sample of circular observations from vM distribution with unknown mean directions μ_i's and unknown homogeneous concentration parameters κ, that is $\theta_i \sim vM(\mu_i, \kappa)$ for $i = 1, ..., n$. Letting X_i be the vector of linear explanatory variables for the i^{th} case, firstly, Gould (1969) developed a regression model in which the structure

$$\mu_i = \mu_0 + \sum \beta_i X_i, \ i = 1, ...n \tag{2.3}$$

was considered for the mean direction, where, μ_0 is an unknown constant, and β_i is the vector of unknown regression coefficients. The maximum likelihood estimates of the model parameters were obtained by using some iterative methods. The likelihood function of this model has with a vengeance many high peaks, which lead to some problems in defining maximum likelihood estimators. Laycock (1975) also examined this model and highlighted that the maximum likelihood estimations are equal to least squares estimations for this model. As an alternative to Gould's model, a different approach was proposed by Johnson and Wehrly (1978) for case of only one linear covariate by using the joint distribution of the circular response variable and linear explanatory variable. They proposed the structure

$$\mu_i = \mu_0 + 2\pi F(x_i), \ i = 1, ...n \tag{2.4}$$

for the mean direction, where $F(x)$ is an exactly determined cumulative distribution function. This alternative approach was extended by Fisher and Lee (1992) to multiple linear covariates by using a general link function. The structure

$$\mu_i = \mu_0 + g(\beta^T X_i),\ i = 1, ...n, \tag{2.5}$$

was suggested by them, where β is vector of regression coefficients, X_i is the vector of predictors for the i^{th} case and g is a general link function. They mapped the real line to the unit circle by using the link function and just monotone link function was considered. The function g should range from $-\pi$ to π with $g(0) = 0$. Some useful link functions are suggested such as $g(x) = 2\pi F(x)$ where F(x) is cumulative distribution function and $g(x) = 2tan^{-1}(x)$. Maximum likelihood estimates of model parameters can be obtained by using some iterative methods.

In general, vM distribution becomes less appealing for more general models in which asymmetry and multi-modality is of concern. A mixture of vM distributions or alternative circular distributions like the wrapped distributions or the projected normal distribution would be a good solution to handle asymmetry and multi-modality.

2.1.1.2 Based on Wrapped Distributions

Another popular approach used in circular-linear regression models is the use of a wrapped distribution. In this approach, a real-line random variable is wrapped on the circumference of a unit circle to obtain a circular random variable. Let X be a real-line a random variable and wrap X on the circumference on a unit circle. Then corresponding circular random variable θ would be $\theta = X(mod 2\pi)$. Applying this idea advanced the utilization of wrapped distributions to induce the dependency of circular response on linear covariates.

Letting g denote the probability density function of real line variable X, then the corresponding circular response follows a wrapped distribution with the following density (Mardia and Jupp, 1999),

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2\pi k). \tag{2.6}$$

The properties of wrapped distributions are very well detailed in Mardia and Jupp (1999).

Fisher and Lee (1994) built a circular regression model under the assumption of wrapped normal distribution. They proposed the Expectation Maximization (EM) algorithm to estimate the model parameters. Since E-step of EM algorithm in problem involves infinite sums, the method is computationally inefficient.

Ravindran and Ghosh (2011) proposed data augmentation method using slice sampling to overcome computational difficulties in the EM algorithm. Their method is flexible and computationally efficient for a wide class of wrapped distributions. In their work, Bayesian method was used to perform regression on circular data. They employed Gibbs sampling method to infer over the posterior distributions.

A method of non-linear transformation of the response space into a new space was proposed by Snelson et al. (2003). They used Gaussian Process algorithm to model the response being in new space. The observation space was wrapped into the latent space by using a non-linear monotonic function.

Sikaroudi and Park (2019) work with a wrapped normal distribution for modeling a circular response. They modeled a circular-linear regression with a mixture of linear-linear regression models. The statistical inference algorithm and the identifiability were well established for their model.

2.1.1.3 Based on Projected Normal Distribution

Projected normal distribution (also called as offset normal distribution by Mardia (1972)) is another applicable distribution which is used to propose an alternative approach for modeling a circular response with linear covariates. This probability distribution is symmetric and unimodal like vM distribution. Basically, in this alternative approach, a linear function of explanatory variables are employed to explain the variation of the mean vector of the bivariate normal distribution.

Presnell et al. (1998) assumed circular response variable being on a unit sphere and proposed the spherically projected multivariate linear model based on projected normal distribution. They used iterative methods such as the Newton–Raphson and the EM algorithms to obtain the maximum likelihood estimation of their regression model parameters. A major drawback of their model is that the approximation of the asymp-

totic variance for the estimated regression parameters can be misleading when the actual variance is high.

Expanding Presnell et al. (1998), Nuñez-Antonio and Gutiérrez-Peña (2005) and Nuñez-Antonio et al. (2011) employed Bayesian approach to perform analysis on projected normal distribution. Gibbs sampling method was used in order to sample from the posterior distributions after the appropriate latent variable is introduced.

More recently Wang and Gelfand (2013) employed a general projected normal distribution which can be asymmetric and bimodal to implement more flexible circular regression modeling with linear covariates. They worked in a fully Bayesian framework. In 2014, they continued their work in 2013, and proposed a projected Gaussian spacial process which is made based on the idea of the projection of a bivariate spacial process on a circle. They introduced a latent variable and used Bayesian approach in their process. They derived the full conditional distributions to perform Gibbs sampling method for having inference of the posterior distributions.

Hernandez-Stumpfhauser et al. (2016) included the factors into the mean vector and proposed a mixed effect model based on the projected normal distribution. They used a Hamiltonian Monte Carlo sampler to estimate regression parameters.

In general, working on the regression models using projected normal distribution has some difficulties as these regression models are usually very complicated. A major drawback of using Bayesian approach on regression models based on the projected normal distribution is that there is a latent variable and it follows the general exponential family formulation which does not have any direct method of sampling. Therefore, the computation time for this method is considerably high.

2.1.1.4 Based on Non-Parametric Regression Method

Another alternative approach is to use a non-parametric regression method for analyzing circular-linear regression models. Marzio et al. (2013) proposed a non-parametric regression model for this problem. In their model, non-parametric smoothing method was used to estimate a conditional mean response using local averaging. The aim of the approach is to find an unknown regression function $\mu(.)$ that minimizes the

circular risk function,

$$E[1 - cos(\theta - \mu(X))|X = x], \tag{2.7}$$

where, $\theta \in [-\pi, \pi]$ is the circular response variable and X is a predictor defined on the real line.

2.1.2 Linear-Circular Regression Models

Linear-circular regression models are a statistical technique used to analyze relationships between some circular predictor variables and a linear response variable.

In this type of regression models, the predictor variable is typically represented using trigonometric functions (e.g., sine and cosine) to take into account its circular nature. There are several approaches to perform linear-circular regression, and discussion on the development of these regression models began with Mardia (1976). A natural linear-circular regression model is

$$X_i|\theta_i \sim N(\alpha + \beta_1 \cos\theta_i + \beta_2 \sin\theta_i, \sigma^2),\ i = 1, ...n \tag{2.8}$$

as proposed by Mardia (1976), where X is linear response variable, α and $\beta_i, i = 1, 2$ are the regression coefficient, θ is the circular independent variable and σ^2 is scale parameter.

Johnson and Wehrly (1978) gave a slight extension of the model in 2.8 and predicted the mean of a linear response variable from a vector of circular covariates. They proposed parametric models for linear-circular dependency based on maximum entropy conditional distribution. Then, SenGupta (2004) generalized the method of Johnson and Wehrly (1978) to encompass possibly asymmetric directional data on manifolds.

More recently, asymmetric linear–circular multivariate regression models involving a simple cosine function was proposed by SenGupta and Ugwuowo (2006). They employed the methods of linear and non-linear least squares to estimate model parameters. They applied them to several real-life environmental data sets.

Bhattacharya and SenGupta (2009) proposed a semi-parametric Bayesian hierarchical framework for handling all forms of uncertainty arising in a linear–circular data set depending on the basis of one of the parametric models in SenGupta and Ugwuowo (2006).

Kim and SenGupta (2015) used the model given by

$$X_i = a + b\cos(\theta_i - \mu) + \epsilon_i,\ i = 1, ...n \tag{2.9}$$

for the regression of a linear variable on a circular variable, where a and b are the regression coefficients, $\mu \in (0, \pi]$ is the mean direction of the circular independent variable, θ. ϵ is error term which have 0 mean and a variance σ^2 and X is a linear response variable.

2.1.3 Circular-Circular Regression Models

Circular-circular regression models are a type of statistical analysis used when both the dependent and independent variables are circular. These models are designed to handle circularity and provide a accurate way to model the relationship between two circular variables. They take into consideration the periodic nature of the circular data and incorporate circular statistics to estimate the parameters.

Sarma and Jammalamadaka (1993) introduced a regression model for two circular random variables θ and ϕ in terms of the conditional expectation of the vector $e^{i\theta}$ given ϕ, namely

$$E[e^{i\theta}|\phi] = \rho(\phi)e^{i\mu(\phi)} = g_1(\phi) + ig_2(\phi), \tag{2.10}$$

where $i = \sqrt{-1}$, $\mu(\phi)$ is the conditional mean direction of θ given ϕ. $\rho(\phi)$ is the conditional concentration parameter for a certain periodic functions $g_1(\phi)$ and $g_2(\phi)$, where $0 \leq \rho(\phi) \leq 1$. Since the estimation of $g_1(\phi)$ and $g_2(\phi)$ is difficult, they are expressed in terms of their Fourier series expansions.

Rivest (1997) proposed a circular-circular regression model in which the circular re-

sponse variable is regressed on a circular predictor variable by decentring the dependent variable.

Downs and Mardia (2002) used tangent link function for regression model involving two circular random variable, such as θ and ϕ. They expressed the regression curve as a form of the Möbius transformtaion or tangent function as following

$$\tan\frac{1}{2}(\theta-\mu_\theta)=\beta\tan\frac{1}{2}(\phi-\mu_\phi), \tag{2.11}$$

where θ and ϕ are circular random response and circular fixed covariate with mean directions μ_θ and μ_ϕ, respectively. β is a slope parameter in the closed interval $[-1,1]$. This mapping defines a one-to-one correspondence between the independent circular variable and the mean of the dependent circular variable. It has a unique solution. They assumed that θ given ϕ has the von Mises distribution with mean direction which is function of ϕ and concentration parameter $\kappa\geq 0$.

In cases when two circular variables θ and ϕ are linearly correlated, Hussin, et al. (2004) proposed the simple linear regression model for predicting a dependent circular random variable on the basis of an independent circular random variable in the following form

$$\theta_i=\alpha+\beta\phi_i+\epsilon_i(mod2\pi)\,,i=1,...,n \tag{2.12}$$

where ϵ_i is the circular random error having a vM distribution with circular mean 0 and concentration parameter κ. This approach is very simple and easy to implement, but it has some major drawbacks. One of the drawbacks of this method is that it does not give the same prediction for θ and $\theta+2\pi$ if the regression coefficient is not an integer. Moreover, it has difficulty due to the dependence on the choice of the angle 0. In other words, it gives different non-equivalent regression functions for different orientations of 0 degrees of the independent circular random variable.

Kato et al. (2008) proposed a circular–circular regression model depending on the Möbius transformation method. They expressed the regression curve as a form of Möbius circle transformation. For a circular random response variable θ and circular

random covariate ϕ, they proposed the regression curve to be defined by

$$\theta_i = \beta_0 \frac{\phi_i + \beta_1}{1 + \bar{\beta}_1 \phi_i} \, , i = 1, ..., n \tag{2.13}$$

where β_0 and β_1 are complex parameters with $\beta_0 \in \Omega$, and $\beta_1 \in \mathbb{C}$, where $\mathbb{C}$ is the complex plane. They worked under the assumption of wrapped Cauchy distribution. They presented some desirable properties of the model, including estimation and testing procedures.

SenGupta et al. (2013) presented a generalization of the model introduced by Downs and Mardia (2002) as following

$$\theta_i = \mu_\theta + 2 \arctan\{\alpha + \beta \tan(\frac{\phi_i - \mu_\phi}{2})\} + \epsilon_i \, , i = 1, ..., n \tag{2.14}$$

where α and β are intercept and slope parameters, respectively. μ_θ and μ_ϕ are the mean directions of θ and ϕ. The circular random error, ϵ follows a circular distribution with 0 mean direction. They assumed two different error distributions, the vM error distribution and the asymmetric generalized vM error distribution.

Abuzaid and Allahham (2015) proposed a simple linear model for circular variables with a wrapped Cauchy error due to its desirable characteristics of having a heavy tailed. They employed an iterative procedure to obtain the maximum likelihood estimates of model parameters. Bootstrapping methods were used to obtain the standard error of the estimates as well as their confidence intervals. They investigated the unbiasness, consistency and robustness of estimates via a simulation study.

Jha and Biswas (2017) extended the idea of Kato et al. (2008) model which is working with only one circular covariate and proposed a multiple circular-circular regression model.

Kim and SenGupta (2017) proposed a fully model-based approach in which the functional relationships between a multivariate circular-dependent variable and several circular covariates can be investigated. This approach enable to inference regarding all model parameters and related prediction. They proposed two multiple circular regression models. The first one is the arc-tangent-sine link model which is defined as

following

$$\theta_i = \mu_\theta + 2\arctan\{\alpha + \sum_{j=1}^{k} \beta_j \sin(\frac{\phi_{ji} - \mu_{\phi_j}}{2})\} + \epsilon_i \, , i = 1, ..., n \tag{2.15}$$

where ϵ circular random error term following a circular distribution with zero mean direction. $\alpha \in \mathbb{R}$ and $\beta_j \in \mathbb{R}$ are the intercept and slope parameters, respectively for $j = 1, ..., k$, k is the number of circular independent variable. μ_θ is the centering constant of θ and μ_{ϕ_j} is the centering constant of ϕ_j. Their second model is the arc-tangent-tangent link model which is an extension of the model introduced in Kim and SenGupta (2015) and defined as following

$$\theta_i = \mu_\theta + 2\arctan\{\alpha + \sum_{j=1}^{k} \beta_j \tan(\frac{\phi_{ji} - \mu_{\phi_j}}{2})\} + \epsilon_i \, , i = 1, ..., n \tag{2.16}$$

where all terms have the same interpretation as in 2.15. Firstly, they focused on an univariate circular-dependent variable ,and then presented a straightforward extension to the case with a multivariate-dependent circular variable with some circular covariates.

2.1.4 Circular-(Circular, Linear) Regression Models

This section combines circular-circular and circular-linear regression models to obtain circular-(circular, linear) regression models. Combining these models imply the development of a statistical framework that can handle relationships between circular response variable and both circular and linear covariates. The process of combining these models involves extending existing statistical methods. It is important because real-world data can involve a mix of circular and linear variables. Suppose one observes $(\theta_i, \phi_{hi}, x_{ki})$, $i = 1, ...n$, $h = 1, ..., p$, $k = 1, ..., q$, where p and q are the number of circular and linear covariates, respectively. $\theta_i \in [0, 2\pi]$ and $\phi_{hi} \in [0, 2\pi]$ are circular, and $x_{ki} \in (-\infty, \infty)$ are linear measurements for $i = 1, ..., n$. For the regression of a circular random variable on some circular and linear variables, the models in Equation 2.5, 2.11 and 2.15 can be combined as follows;

$$\theta_i = \mu_\theta + 2\arctan\{\sum_{h=1}^{p} \beta_h \tan(\frac{\phi_{hi} - \mu_{\phi_h}}{2}) + \sum_{k=1}^{q} \beta_k x_{ki})\} + \epsilon_i, \tag{2.17}$$

$$\theta_i = \mu_\theta + 2\arctan\{\sum_{h=1}^{p} \beta_h \sin(\frac{\phi_{hi} - \mu_{\phi_h}}{2}) + \sum_{k=1}^{q} \beta_k x_{ki})\} + \epsilon_i, \tag{2.18}$$

where $i = 1, ..., n$, $h = 1, ..., p$, $k = 1, ..., q$.

In this thesis, the regression model given in Equation 2.18 is used as a regression model of a circular random variable on some circular and linear variables. α parameter is removed from the model given in Equation 2.15 so that the expected value of θ is the mean direction μ_θ when the values of ϕ_h's are at their mean direction μ_{ϕ_h}'s and x_k's are equal to zero. Note that instead of focusing on circular-linear and circular-circular regression models separately, we focus on circular-(circular,linear) regression models.

2.2 Variable Selection in Circular Regression Models

In this section, variable selection methods for circular regression models are reviewed in detail. Variable selection in regression models is the process of choosing the significant and important predictors to include in the model while excluding insignificant or redundant ones. The goal of variable selection is to create a more parsimonious and interpretable model that can accurately predict the dependent variable without unnecessary complexity. Deciding the best parsimonious model is a crucial step in the analysis of data since the subsequent inference is based on the best parsimonious model is found in traditional statistical data analysis.

Circular literature for obtaining parsimonious model is given below in a chronological order. Lund (1999) used usual statistical tests using the asymptotic variances of the parameter estimates and a bias-corrected Akaike information criterion statistic under the assumption of vM errors. D'Elia (2001) used a test statistic which is based on the likelihood ratio criterion to compare some nested models and then by using AIC, they preferred the model which has fewer parameters. Downs and Mardia (2002) employed some test statistics such as Wald score and F statistic to select

the best model. Carnicero et. al. (2008) used BIC to select the number of terms to include in their mixture model. Ravindran and Ghosh (2011) defined a statistical tool called as Circular Predictive Discrepancy to decide the best model by using the decision theoretic framework of Gelfand and Ghosh (1998). 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. SenGupta et al. (2015) employed pseudo Bayes factor for the purpose of model comparison. 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. 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 was 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. SenGupta and Kim (2016) employed likelihood ratio tests for comparing some regression models. They also used usual hypothesis testing procedure for testing the regression paramaters under the asymptotic normality assumption for the parameter estimators. Again Kim and SenGupta (2017) employed a circular mean-square error (CMSE) to compare some regression models and used usual hypothesis testing procedure such as Wald test for testing the regression paramaters. Camli and Kalaylioglu (2021) employed the prediction error based model selection criteria such as CPD_1 and CPD_2 to compare the performance of circular random effects models. They also used plug-in estimators of CPD_1 and CPD_2 for comparing the prediction abilities of the models. Sikaroudi and Park (2019) define the mean circular error using sine function and used it to select the better model by measuring the errors of models.

There have been a various of variable/model selection methods in the literature from both frequentist and Bayesian perspectives for circular regression models. On the other hand, there is no any literature on Bayesian penalization techniques such as Bayesian lasso for circular regression models to improve the prediction ability of

the model and to reduce model complexity and the risk of overfitting by giving a parsimonious model.

CHAPTER 3

SHRINKAGE-BASED VARIABLE SELECTION IN CIRCULAR REGRESSION MODELS

In this thesis, we consider new approaches for variable selection in circular regression models. Inspired by the Bayesian lasso technique used in Euclidean space modeling, these approaches create a set of ways to choose variables in circular regression models. With the growing need for more accurate and adaptable circular data modeling, the methods introduced here are a crucial contribution that could change how variable selection is done in circular regression analysis.

One major achievement is the introduction of a new prior distribution for the tuning parameter λ in the Bayeasin lasso. This prior distribution can be used effectively in both Euclidean and circular regression settings. Since λ is important in various modeling areas, its successful incorporation is a significant improvement in statistical practice. As a result, this contribution has the potential to get accurate and strong regression analysis.

Based on these developments, this chapter delve into different variations of the lasso regression technique, including the lasso introduced by Tibshirani (1996), the Bayesian lasso by Park and Casella (2008) and the Bayesian adaptive lasso of Leng et al. (2014). The chapter also discusses various methods for specifying the tuning parameter for both the lasso and Bayesian lasso, including the use of different priors. Furthermore, the chapter introduces a new empirical Bayes (EB) hyper-prior to estimate the tuning parameter by incorporating observed data. In the last part, which is the chief part of this section, the chapter focuses on developing a Bayesian lasso types adaptation in circular regression models.

3.1 Review of the lasso

In this section, firstly, we briefly explain the lasso in Euclidean space. Then we review the Bayesian lasso and give the advantages of it.

3.1.1 Classical Lasso

Assume that we have a linear covariate vector $\mathbf{X} = (X_1, X_2, ..., X_p)^T$, and want to predict a real-valued linear response $\mathbf{Y}$. Consider the following linear regression model:

$$\mathbf{Y} = \beta_0 + \sum_{j=1}^{p} X_j \beta_j + \boldsymbol{\epsilon}, \tag{3.1}$$

where the $\beta_j' s$ denote unknown regression parameters which can be estimated and error terms, $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I)$. The most popular estimation method is ordinary least squares (OLS) for parameter estimation in regression problems with linear variables.

When we use the OLS estimates, two challenges can take place: *i.* prediction accuracy: We would like our model to accurately predict future data. The least squares estimates often have low bias at the cost of large variance. Prediction accuracy can often be improved by shrinking the regression coefficients. Shrinkage sacrifices some bias to reduce the variance of the predicted value and hence may improve the overall prediction. *ii.* Interpretation: OLS cannot distinguish variables with little or no influence. These variables cause in applications distraction from the significant regressors. Practitioner's intention is to select only a subset of all variables which is assumed to be significant.

Tibshirani (1996) proposed a new technique, called the lasso (least absolute shrinkage and selection operator). Lasso is a regularization technique used in linear regression. It is primarily employed to prevent over-fitting and enhance model generalization. Lasso introduces a penalty term to the regression objective function, encouraging the model to favor simpler and more interpretable models by driving some of the coefficient values to exactly zero.

Letting $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p)^T$, lasso coefficients are solutions to the l_1 optimization problem and the lasso estimate $\hat{\beta}^{lasso}$ is defined by

$$\hat{\beta}^{lasso} = \arg min\{||\mathbf{Y} - \mathbf{X}\beta||_2^2\} \ \textit{s.t.} \ ||\beta||_1 = \sum_j |\beta_j| \leq t \tag{3.2}$$

or equivalently (Murray and Wright, 1981)

$$\hat{\beta}^{lasso} = \arg min\{||\mathbf{Y} - \mathbf{X}\beta||_2^2 + \lambda||\beta||_1\}, \tag{3.3}$$

where t ≥ 0 and $\lambda \geq 0$ are tuning parameters. The lasso penalty contains an absolute value, thus, the objective function is not differentiable. Therefore, in general, the lasso solution lacks a closed form. This requires the implementation of an optimization algorithm to find the minimizing solution. Since both the objective function and the constraint are convex functions, the lasso estimate can be solved by standard convex optimization techniques. Tibshirani (1996) described some efficient and stable algorithms for the solution of this problem and so the lasso has become a widely used alternative to ordinary least squares for parameter estimation in regression problems.

The tuning parameter λ controls the strength of penalty, which shrinks each β_j towards zero. For λ close to zero, the lasso estimate is close to the OLS. Whereas large enough λ will set some coefficients exactly equal to 0. So the lasso will perform variable selection. As λ increases, more coefficients are set to zero (less variables are selected), and among the nonzero coefficients, more shrinkage is employed. To arrive at a final lasso estimator of β, the lasso penalty parameter λ needs to be chosen. Three different methods can be used for the estimation of the lasso parameter λ: *i.*cross-validation, *ii.* generalized cross-validation and *iii.* an analytical unbiased estimate of risk (Tibshirani, 1996).

Here are some advantages of using lasso:

- Variable selection: Lasso tends to set the coefficients of less relevant variables to zero, effectively selecting the most important variables in the dataset. This variable selection property can be beneficial when dealing with high-

dimensional data, reducing the model's complexity and improving interpretability.

- Improved model generalization: By penalizing large coefficient values, lasso helps prevent over-fitting, leading to better performance on unseen data. Regularization techniques like lasso are particularly useful when dealing with limited training data.
- Interpretable models: The variable selection nature of lasso can lead to simpler models with fewer predictors, making the model more interpretable and easier to understand.
- Dealing with multicollinearity: Lasso can handle multicollinearity (high correlation between predictors) better than ordinary least squares regression. It will tend to pick one of the correlated variables and set the others to zero.
- Regularization strength control: Lasso has a tuning parameter that allows to control the strength of regularization. By using this parameter, one can adjust the amount of shrinkage applied to the coefficients and balance the trade-off between bias and variance in the model.
- The lasso can also accommodate models other than the linear models described here. It is possible to use lasso estimators in generalized linear models, including logistic regression, multi-class logistic regression, Cox proportional hazard models, and nonlinear regressions (Tateishi et al., 2010).

However, it's essential to note that lasso may not perform well in situations where the number of predictors is much larger than the number of observations or when there are groups of highly correlated predictors (Zou and Hastie, 2005). In such cases, other regularization techniques like Elastic Net (Zou and Hastie, 2005) or Ridge Regression (Hoerl and Kennard, 1970) might be more suitable.

3.1.2 Bayesian Lasso

In this section we review the Bayesian lasso and provide all its details. The Bayesian lasso extends the lasso regression by incorporating a prior distribution on the regres-

sion coefficients. Tibshirani (1996) stated that the lasso estimates can be obtained by calculating the mode of the conditional posterior distribution of the regression coefficients in a Bayesian model in which each coefficient is assigned the same DE prior. The DE distribution is given by the probability density function:

$$f(x|\mu, b) = \frac{1}{2b} e^{-|x-\mu|/b}, \tag{3.4}$$

with mean μ and variance $2b^2$, where $x, \mu \in \mathbb{R}$ and $b > 0$.

When a DE prior is assumed for regression coefficients, β with a density function with location $\mu = 0$ and scale $\frac{1}{\lambda}$

$$\pi(\beta|\lambda) = \frac{\lambda}{2} e^{-\lambda|\beta|}, \tag{3.5}$$

where $\beta \in \mathbb{R}, \lambda > 0$, the joint posterior distribution of regression coefficients, β is proportional to the following quantity

$$f(\beta|\mathbf{X}, \sigma^2, \lambda) \propto \exp(-\frac{1}{\sigma^2}||\mathbf{Y} - \mathbf{X}\beta||_2^2 + \lambda||\beta||_1). \tag{3.6}$$

The equivalency between two objectives in Equation 3.3 and in Equation 3.6 can be seen easily.

If a DE distribution prior is directly employed, developing Gibbs samplers for Bayesian lasso involves some computational challenges as the complexity in computation increases. On the other hand, one of the advantages of the DE distribution is that it can be expressed as a scale mixture of normal distributions with independent exponentially distributed variances (Andrews and Mallows, 1974) as follows

$$
\begin{aligned}
P(\beta|\sigma^2) &= \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}} \\
&= \prod_{j=1}^{p} \int \frac{1}{\sqrt{2\pi\sigma^2\tau_j^2}} e^{\frac{-\lambda|\beta_j|^2}{2\sigma^2\tau_j^2}} \frac{\lambda^2}{2} e^{-\lambda^2\tau_j^2/2} d\tau_j^2 \\
&= \prod_{j=1}^{p} \int N(\beta_j|\sigma^2, \tau_j^2) Exp(\tau^2|\frac{\lambda^2}{2}) d\tau_j^2
\end{aligned}
\tag{3.7}
$$

Here, N(.) is normal distribution and Exp(.) is exponential distribution.

Due to this Bayesian interpretation of the lasso, there have been a variety of Bayesian variations of the lasso in the past couple of years. Different approaches based on representing the DE distribution as scale mixture of normal distributions with independent exponentially distributed variances have been proposed (Figueiredo, 2003; Bae and Mallick, 2004; Yuan and Lin, 2005). Park and Casella (2008) provided the first explicit treatment of Bayesian lasso regression. They introduced a fully Bayesian lasso. Then, Hans (2009) proposed new aspects of the broader Bayesian treatment of lasso regression. They provided a direct characterization of the regression coefficients's posterior distribution. They used the posterior mean as point estimation for regression coefficients. Chen et al. (2011) proposed a new, fully hierarchical, Bayesian version of the Lasso. They employed a reversible-jump Markov Chain Monte Carlo (MCMC) algorithm to obtain the Bayesian Lasso estimate. Mallick and Yi (2014) proposed a different hierarchical formulation of Bayesian lasso. Their formulation is based on representing of the DE distribution as the scale mixture of uniform distribution. They consider a fully Bayesian lasso and propose a new Gibbs sampler with tractable full conditional posterior distributions.

In the literature, although there have been various version of Bayesian lasso, in this thesis we focus on the Bayesian lasso of Park and Casella (2008). They considered a fully Bayesian lasso using a conditional DE prior as follows

$$
P(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}}. \tag{3.8}
$$

They stated that conditioning on σ^2 is important since it guarantees a unimodal full

posterior for β. If the full posterior is not unimodal, it could slow the convergence of the sampling process and the resulting point estimates might not be meaningful (Park and Casella, 2008).

They proposed the following hierarchical model by using the alternative representation of Equation 3.8 given in Equation 3.7,

$$1^{st}\ level \quad \mathbf{Y}|\mu, \mathbf{X}, \beta, \sigma^2 \sim N(\mu + \mathbf{X}\beta, \sigma^2 I) \tag{3.9}$$

$$2^{nd}\ level \quad \beta|\tau_1^2, ..., \tau_p^2, \sigma^2 \sim N(0, \sigma^2 \mathbf{D}_\tau),\ \mathbf{D}_\tau = diag(\tau_1^2, ..., \tau_p^2) \tag{3.10}$$

$$3^{rd}\ level \quad \sigma^2, \tau_1^2, ..., \tau_p^2 \sim \pi(\sigma^2)d\sigma^2 \prod_{j=1}^{p} \frac{\lambda^2}{2} e^{-\lambda^2 \tau_j^2/2} d\tau_j^2,\ \tau_1^2, ..., \tau_p^2 > 0 \tag{3.11}$$

where, the first level denotes the likelihood, the second level is the prior distribution for regression coefficients, β, and finally, the third level is the prior for scale parameters. Here, μ is the overall mean. It can be given a flat prior and subsequently integrated out. Based on this hierarchical model, a Gibbs sampler can be implemented with the following full conditional distributions.

$$\mu \sim N(\bar{y}, \sigma^2/n), \tag{3.12}$$

$$\beta \sim N(\mathbf{A^{-1}X^T\tilde{y}}, \sigma^2 \mathbf{A^{-1}}),\ \mathbf{A} = \mathbf{X^TX} + \mathbf{D_\tau^{-1}}, \tag{3.13}$$

$$\sigma^2 \sim Inv - Gamma((n-1)/2 + p/2, (\tilde{\mathbf{y}} - \mathbf{X}\beta)^{\mathbf{T}}(\tilde{\mathbf{y}} - \mathbf{X}\beta)/2 + \beta^{\mathbf{T}}\mathbf{D}_\tau^{-\mathbf{1}}\beta/2), \tag{3.14}$$

$$\frac{1}{\tau_j^2} \sim Inv - Gauss(\sqrt{\frac{\lambda^2\sigma^2}{\beta_j^2}}, \lambda^2), \tag{3.15}$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \bar{\mathbf{y}}\mathbf{I}$. They used the median of the posterior distribution as point estimates of the regression coefficients.

The tuning parameter, λ can be considered as a hidden parameter in Bayesian lasso since the first level (likelihood) does not contain it. Although λ in classical lasso directly affects the estimations of β, in Bayesian lasso, it indirectly affects the estimation of β by controlling the variation of local scales τ_j^2. In other words, the tuning

parameter λ in the lasso formulation is often related to the precision of the Laplace prior in Bayesian lasso.

Park and Casella (2008) offered two different methods that address the choice of tuning parameter such as a full Bayesian approach and an empirical Bayes approach. In the empirical Bayes approach, they employed empirical Bayes through marginal maximum likelihood and took advantage of an expectation maximization algorithm to complement the Gibbs sampler. In this thesis, we concentrate on a full Bayesian approach. In the full Bayesian approach, they considered the use of an appropriate hyper-prior like the class of Gamma priors on λ^2 which is more convenient than defining a prior on λ of the form

$$\pi(\lambda^2) = \frac{\delta^r}{\Gamma(r)}(\lambda^2)^{r-1}e^{\delta\lambda^2},\ \lambda^2 > 0,\ r > 0,\ \delta > 0. \tag{3.16}$$

The full conditional distribution of λ^2 is gamma with the shape parameter p + r and rate parameter $\sum_{j=1}^{p} \tau_j^2/2+\delta$. They point out that λ^2 can easily join the other parameters and the full conditional distribution of λ^2 does not affect those of the other parameters. Moreover, in the Bayesian lasso literature, half-Cauchy distribution which is known as a robust alternative and a popular prior distribution in the Bayesian literature (Gelman, 2006; Polson and Scott, 2012; Mulder and Pericchi, 2018) is considered as the prior distribution for the tuning parameter (Van Erp et al., 2019).

The advantage of using Bayesian lasso over the classical lasso lies in the Bayesian framework's ability to provide probabilistic estimates and uncertainty quantification. Here are some advantages of the Bayesian lasso:

- Unlike the classical lasso, which only provides point estimates of the regression coefficients, Bayesian lasso provides posterior distributions for the coefficients. These distributions capture the uncertainty associated with the estimates, which can be valuable in decision-making and sensitivity analysis.

- The lasso is known for its ability to perform automatic variable selection, meaning it can effectively shrink some regression coefficients to exactly zero, effectively excluding those variables from the model. Bayesian lasso inherits this

property, and by providing probabilistic estimates, it can offer better insights into which variables are more likely to be relevant or irrelevant.

- In the Bayesian framework, one can introduce prior information about the model coefficients. This is particularly useful when one have some prior knowledge about the effect sizes or relationships between variables. The prior allows us to regularize the coefficients and can lead to more robust and informative parameter estimates.

- Bayesian lasso facilitates model averaging by combining information from multiple models, each with different penalization strengths. This helps in dealing with model uncertainty and stabilizes the coefficient estimates when dealing with multicollinearity or high-dimensional data.

- The Bayesian lasso allows us to use different penalty functions, not just the L_1 (absolute value) penalty used in the classical Lasso. This flexibility enables the use of alternative priors, which can better suit the data and lead to improved model performance.

- Bayesian lasso is well-suited for high-dimensional datasets, where the number of predictors is much larger than the number of observations. The Bayesian framework's ability to provide uncertainty estimates helps in dealing with the challenges of over-fitting in such situations.

- Bayesian lasso is analytically simpler and is also easy to implement. Penalization and shrinkage can be achieved naturally in fully Bayesian framework thanks to the prior distributions of regression coefficients and the tuning parameter.

- All model parameters and the tuning parameter can be estimated in one step. Also, their credible intervals (CIs) are easily obtained from MCMC sampling procedure.

Overall, Bayesian lasso is a powerful method for sparse regression and variable selection, combining the advantages of Bayesian modeling with the variable selection capabilities of lasso.

3.1.3 Bayesian Adaptive Lasso

In this section, we briefly review the Bayesian adaptive lasso (Leng et al., 2014). In that study, the following adaptive lasso estimator is considered:

$$\hat{\beta}^{lasso} = \min_{\beta}((y - X\beta)^T(y - X\beta) + \sum_{j=1}^{p} \lambda_j|\beta_j|), \tag{3.17}$$

where different penalty parameters are employed for the regression coefficients. Here λ_j's are used instead of λ, everything else is the same as in the work of Park and Casella (2008).

3.1.4 Specification of the tuning parameter

There is an extensive literature on tuning parameter specification. In the frequentist framework, the methods can be divided into the following broad categories: *i.* resampling based procedures such as cross-validation, generalized cross-validation, modified cross-validation and the bootstrap (Tibshirani, 1996; Hall et al., 2009; Yu and Feng, 2014), *ii.* information criterion based approaches such as AIC, BIC, and generalized information criterion (Fan and Li, 2001; Wang et al., 2007; Wang et al., 2009; Tateishi et al., 2010; Bühlmann and Van De Geer, 2011; Flynn et al., 2013; , Fan and Tang, 2013; Inoue and Nagata, 2016), *iii.* model metrics based methods such as the residual sum of squares or Mallows's Cp-type selection criterion (Efron et al., 2014; Inoue and Nagata, 2016), *iv.* an analytical unbiased estimate of risk (Stein's unbiased risk estimation) (Tibshirani, 1996).

In the Bayesian framework in which uncertainty in the tuning parameter is accounted for by assigning hyper-prior density apriori, the approaches include *i.* an EB approach through marginal maximum likelihood (Park and Casella, 2008), *ii.* Bayes factor (Lykou and Ntzoufras, 2013), *iii.* selecting an appropriate hyper-prior Park and Casella, 2008; Hans, 2009; Kyung et al., 2010; Lykou and Ntzoufras, 2013; Mallick and Yi, 2014; Van Erp et al., 2019), *iv.* information criterion based approaches such as DIC (Kawano et al., 2015). The focus of this thesis is selecting an appropriate hyper-prior for the tuning parameter. A suitable hyper-prior can be placed on λ or λ^2.

For instance, Park and Casella (2008) considered the class of gamma hyper-priors on λ^2 which is given by:

$$\pi(\lambda^2) = \frac{\delta^r}{\Gamma(r)}(\lambda^2)^{r-1}e^{-\delta\lambda^2},\ \lambda^2 > 0,\ r > 0,\ \delta > 0, \tag{3.18}$$

where, r and δ are shape and rate parameters, respectively. Since the choice of hyper-prior distribution has effect on subsequent inference as shown in Section 4.1.2.2, choosing an appropriate hyper-prior distribution for λ is crucial. Notice that the full conditional distribution of λ^2 depends on $p, r, \sum_{j=1}^{p}\tau_j^2$ and δ where p is fixed and known, τ_j's are estimated from the data during the analysis and hyper-hyper parameters, r and δ are set by the analyst. There are currently two methods to set r and δ apriori. One of them is given by Park and Casella (2008) where r and δ are determined such that hyper-prior distribution has a high probability near the MLE of λ and relatively flat otherwise. However finding the MLE of λ is difficult as it is based on a computationally very intensive approach and the rate of convergence heavily relies on the initial. In addition, configuration the flatness depends on the subjective choice of the analyst for r and δ. Since the selection of the tuning parameter or the hyper-hyper parameters (r and δ) controls the entire procedure, their selection is desired to be free of researcher's subjective effect and effect of initial values. The other method is based on correlation measures such as benchmark and threshold correlations given by Lykou and Nzourfras (2013). Their approach gives similar inclusion probabilities for significant covariates to those of a non-informative prior (such as $Ga(0.01, 0.01)$).

3.2 An Empirical Bayes Type Hyper-Prior for λ

Desired properties of hyper-prior for λ are good mixing in MCMC samplers, identifiability of model parameters, and robustness of posterior inference. Otherwise problems of non-identifiability and of sensitivity arise in posterior inference. Prior elicitation based on Empirical Bayes (EB) notion is previously shown to eradicate such problems with hyper-parameters in linear-linear regression. Here we employ the notion of EB to construct hyper-prior for the lasso tuning parameter λ. Our method is based on adopting gamma or half-Cauchy type prior for λ with paramaters estimated

from the empirical data at hand via Gibbs Sampling (GS). GS here approximates the maximum of the marginal likelihood which is otherwise intractable. We call our method EB-GS indicating EB approach led by GS and the prior constructed that way EB-GS prior. The steps of the method are given below in detail.

EB-GS prior is constructed in two distinct steps both of which require GS. Considering a gamma type hyper-prior for λ, namely $Ga(r, \delta)$, the first step is assigning a pair of non-informative priors for r and δ, e.g. $Ga(0.01, 0.01)$, running a Gibbs sampling, and estimating using the modes of the Gibbs samples. Note that, under ergodicity, mode of the Gibbs samples converges to the mode of the marginal posterior density and to the maximum of the marginal likelihood. Second step is employing the estimates resulted from step 1 to construct the hyper-prior distribution of λ i.e., $\lambda \sim Ga(\hat{r}, \hat{\delta})$, where $\hat{r}$ and $\hat{\delta}$ are posterior modes of the posterior distributions of r and δ obtained in step 1, respectively. The steps are given in Table 3.1.

Table 3.1: Algorithm for EB-GS method of prior construction

	Step 1:
1	Assign a non-informative prior for r and δ
2	Implement the Gibbs sampler
3	Obtain posterior mode of r and δ ($\hat{r}$ and $\hat{\delta}$)
	Step 2:
1	Use $\hat{r}$ and $\hat{\delta}$ to assign hyper-prior for λ ($\lambda \sim Ga(\hat{r}, \hat{\delta})$)
2	Implement the Gibbs sampler
3	Obtain posterior mean of parameter of interest (β, σ^2, λ)

3.3 The Proposed Variable Selection Methods

In this section, we first introduce an adaptation of Bayesian lasso in the context of a linear-circular regression model. Second, we focus on the adaptation of Bayesian lasso and Bayesian adaptive lasso in the context of circular-(circular, linear) regression models and introduce some alternative variable selection methods for this type regression models. These alternative methods provide flexibility in handling circular

data and can be useful when the standard circular regression with Bayesian lasso may not be the best option for the specific problem.

3.3.1 Bayesian Lasso in Linear-Circular Regression Models

In linear-circular regression models, cosine representation is a popular choice of modeling (Kim and SenGupta, 2015).

Suppose one observes (y_i, θ_i), where $y_i \in (-\infty, \infty)$ and $\theta_i \in [0, 2\pi]$ (or $\theta_i \in [-\pi, \pi]$) are linear and circular measurements respectively for $i = 1, ..., n$. For the regression of a linear random variable Y on a circular variable θ, we use the model given by Kim and SenGupta (2015), that is

$$Y_i = \alpha + \beta \cos(\theta_i - \mu_\theta) + \epsilon_i, \ i = 1, ..., n \tag{3.19}$$

where, $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are the intercept and slope parameters respectively, $\mu_\theta \in (-\pi, \pi]$ denotes the mean direction of θ, and ϵ_i is error term which follows a linear distribution with zero mean and an unknown variance denoted by σ^2. We reparameterize the model in Equation 3.19 as follows.

$$Y_i = \beta_0 + \beta_1 \cos\theta_i + \beta_2 \sin\theta_i + \epsilon_i, \ i = 1, ..., n, \tag{3.20}$$

where $\beta_0 = \alpha$ is intercept parameter, $\beta_1 = \beta \cos\mu_\theta$ and $\beta_2 = \beta \sin\mu_\theta$ are slope parameters. In this representation, there are two slope parameters for circular covariate. Note that all parameters are linear in the second representation of the model while the first representation of model has a circular parameter μ_θ.

Parameters of the model in Equation 3.20 can be estimated by minimizing the below objective function $Q(.)$, which is L_2 distance, subject to $\sum_{j=1}^{2} |\beta_j| \leq t$

$$Q(\Omega) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 \cos\theta_i - \beta_2 \sin\theta_i)^2, \tag{3.21}$$

where the parameter space is $\Omega = \{\beta_0, \beta_1, \beta_2\}$.

The model in Equation 3.20 can be extended as follows for p circular covariates,

$$Y_i = \beta_0 + \sum_{k=1}^{p} \{\beta_{1k} \cos\theta_{ik} + \beta_{2k} \sin\theta_{ik}\} + \epsilon_i,\ i = 1, ..., n,\ k = 1, ..., p \tag{3.22}$$

where, $\beta_{1k} = \beta_k \cos\mu_{\theta_k}$, $\beta_{2k} = \beta_k \sin\mu_{\theta_k}$ are slope parameters for k^{th} circular covariate.

The parameters of the model in Equation 3.22 can be estimated by minimizing the following objective function $Q(.)$ subject to $\sum_{k=1}^{p}\sum_{j=1}^{2} |\beta_{jk}| \leq t$

$$Q(\Omega) = \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{k=1}^{p} \{\beta_{1k} \cos\theta_{ik} + \beta_{2k} \sin\theta_{ik}\})^2, \tag{3.23}$$

where, the parameter space, $\Omega = \{\beta_0, \beta_{1k}, \beta_{2k}\}_{k=1,...,p}$.

Adaptation of Bayesian lasso for the cosine model is quite straightforward. Combining the model structure we give in Equation 3.22 with the alternative representation for DE of lasso given by Park and Casella (2008) and Andrews and Mallows (1974), we get the following hierarchical structure

$$\begin{aligned}
\mathbf{Y}|\beta_0, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim N(\beta_0 + \mathbf{X}\boldsymbol{\beta}, \sigma^2 I), \\
\boldsymbol{\beta}|\tau_1^2, ..., \tau_{2p}^2, \sigma^2 &\sim N(0, \sigma^2 \mathbf{D}_\tau),\ \mathbf{D}_\tau = diag(\tau_1^2, ..., \tau_{2p}^2), \\
\sigma^2 &\sim Inverse - Gamma(\alpha_\sigma, \beta_\sigma), \\
\tau_j^2 &\sim Exp(\lambda^2/2)\ , j = 1, ..., 2p, \\
\lambda^2 &\sim Ga(r, \delta), \\
\beta_0 &\sim f(\beta_0).
\end{aligned} \tag{3.24}$$

Here, $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)^T$ is $n \times 1$ vector of linear response for $i = 1, ..., n$, $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_{2p-1}}, \boldsymbol{X_{2p}})$ is $n \times (2p)$ matrix of linear covariates, where $\boldsymbol{X_j} = (X_{j1}, X_{j2}, ..., X_{jn})^T$ is $n \times 1$ vector of linear covariate for $j = 1, ..., 2p$ with $X_{1i} = \cos(\theta_{1i})$, $X_{2i} = \sin(\theta_{1i}), ..., X_{(2p-1)i} = \cos(\theta_{pi}), X_{(2p)i} = \sin(\theta_{pi})$, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_{2p-1}, \beta_{2p})^T$ is the vector of regression coefficients, with $\beta_1 = \beta_{11}$,

$\beta_2 = \beta_{21}$,..., $\beta_{2p-1} = \beta_{1p}, \beta_{2p} = \beta_{2p}$, $f(\beta_0)$ is the prior distribution for β_0 which can be selected as an independent, flat prior. Based on this hierarchical model, a Gibbs sampler can be implemented with the following full conditional distributions. Letting $\Omega = \{\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}^2, \lambda^2\}$, where $\boldsymbol{\tau}^2 = (\tau_1^2, \tau_2^2 ..., \tau_{2p-1}^2, \tau_{2p}^2)$, with p denoting the number of circular covariates, full conditional distributions are

$$
\begin{aligned}
&\beta_0 \mid \boldsymbol{Y}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}^2, \lambda^2 \sim N(\bar{y}, \sigma^2/n), \\
&\boldsymbol{\beta} \mid \boldsymbol{Y}, \beta_0, \sigma^2, \boldsymbol{\tau}^2, \lambda^2 \sim N(\mathbf{A^{-1}X^T y}, \sigma^2 \mathbf{A^{-1}}),\ \mathbf{A} = \mathbf{X^T X} + \mathbf{D_\tau^{-1}}, \\
&\sigma^2 \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \lambda^2 \sim Inverse - Gamma((n-1)/2 + p, \\
&\quad ((\boldsymbol{y} - \beta_0 - \boldsymbol{X\beta})^T (\boldsymbol{y} - \beta_0 - \boldsymbol{X\beta}) + \boldsymbol{\beta}^T {\boldsymbol{D_\tau}}^{-1} \boldsymbol{\beta})/2 + \beta_\sigma), \\
&\frac{1}{\tau_j^2} \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \sigma^2, \tau^2, \lambda^2 \sim Inverse - Gaussian(\sqrt{\frac{\lambda^2 \sigma^2}{\beta_j^2}}, \lambda^2), \\
&\lambda^2 \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}^2, \sim Ga(p + r, \sum_{j=1}^{2p} \tau_j^2/2 + \delta).
\end{aligned}
\tag{3.25}
$$

The derivations of the full conditional distributions are provided in Appendix A.1. We used posterior mean as Bayesian lasso estimates. MCMC performed without any convergence problem and the convergence is achieved optimally fast.

Note that for the Bayesian lasso, our posterior density is steep since inverse Hessian matrix, $[-\bar{L}''(\omega_0)]^{-1}$, $\to 0$ as n $\to \infty$ where ω_0 denotes the true value of parameter of interest ω. In other words, the largest eigenvalue of $n/\mathbf{V} \to 0$ as n $\to \infty$ where V is variance matrix. Notice that $\bar{L}''(\omega_0)$ can be obtained easily for the normal distribution. Our posterior density is also smooth since $\bar{L}''(\omega)$ is a continuous function of ω. Since these two conditions are satisfied, posterior estimators for Bayesian lasso in linear-circular regression model have asymptotically multivariate normal distribution. They also have posterior consistency property.

3.3.2 Bayesian Lasso in Circular-(Circular, Linear) Regression Models

In this section, we consider Bayesian lasso in circular-(circular, linear) regression models given in Section 2.1.4. We first consider classical Bayesian lasso as a straightforward implication of Bayesian lasso in circular case (Section 3.3.2.1). In some

cases, adaptive shrinkage is desirable, especially when dealing with high-dimensional data or when certain subsets of variables are more relevant or informative in specific clusters or subgroups. Adaptive shrinkage allows the model to assign different shrinkage levels to different coefficients, effectively allowing for variable selection within each cluster. This can lead to better interpretability and improved predictive performance in scenarios where the data have heterogeneous structures.

To achieve adaptive shrinkage in the hierarchical model, we developed some alternative Bayesian lasso methods. Our methods allow for individual shrinking coefficients for each coefficient. Our new methods are given in Sections 3.3.2.2-3.3.2.5

3.3.2.1 Classical Bayesian Lasso

Assume that $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)^T$ is $n \times 1$ vector of circular response, where $\theta_i \in [0, 2\pi]$ for $i = 1, ..., n$. $\boldsymbol{\phi} = (\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, ..., \boldsymbol{\phi_p})$ is $n \times p$ matrix of circular covariates, where $\boldsymbol{\phi_h} = (\phi_{h1}, \phi_{h2}, ..., \phi_{hn})^T$ is $n \times 1$ vector of circular covariate and $\phi_{hi} \in [0, 2\pi]$ for $h = 1, ..., p$. $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_q})$ is $n \times q$ matrix of linear covariates, where $\boldsymbol{X_k} = (X_{k1}, X_{k2}, ..., X_{kn})^T$ is $n \times 1$ vector of linear covariate and $X_{ki} \in (-\infty, \infty)$ for $k = 1, ..., q$. Let $\boldsymbol{\beta}$ be a $(p + q) \times 1$ column vector of $(\boldsymbol{\beta_\phi}, \boldsymbol{\beta_X})$, where $\boldsymbol{\beta_\phi} = (\beta_{\phi_1}, \beta_{\phi_2}, ..., \beta_{\phi_p})^T$ is $p \times 1$ vector of regression coefficients associated with the circular covariates $\boldsymbol{\phi}_h$ and $\boldsymbol{\beta_X} = (\beta_{X_1}, \beta_{X_2}, ..., \beta_{X_q})^T$ is $q \times 1$ vector of that associated with linear covariates $\boldsymbol{X}_k$. Lastly, let the total number of covariates be denoted by m, i.e. $p + q = m$. We consider von Mises distribution as this is the distribution of choice for circular regression in most applied problems. By using the logic of usual Bayesian lasso, we considered the following hierarchical model,

$$
\begin{aligned}
\boldsymbol{\theta} \mid \mu_\theta, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\beta}, \kappa &\sim vM(\mu_\theta + g((\boldsymbol{\phi}\boldsymbol{\beta_\phi} + \boldsymbol{X}\boldsymbol{\beta_X})), \kappa), \\
\boldsymbol{\beta} \mid \boldsymbol{\tau}^2 &\sim N(0, \mathbf{D}_\tau), \boldsymbol{\tau}^2 = (\tau_1^2, ..., \tau_m^2)\ , \mathbf{D}_\tau = diag(\tau_1^2, ..., \tau_m^2), \\
\kappa &\sim Gamma(a, b), \\
\tau_j^2 &\sim Exp(\lambda^2/2)\ , j = 1, ..., m, \\
\lambda^2 &\sim Gamma(c, d), \\
\mu_\theta &\sim Circ - Unif(0, 2\pi),
\end{aligned}
\tag{3.26}
$$

where μ_θ denotes the mean direction of $\boldsymbol{\theta}$, and κ is concentration parameter. Based on this hierarchical model, full conditional distributions can easily be obtained to implement a Gibbs sampler. The derivations of the full conditional distributions are provided in Appendix A.2.1. Note that this approach is very straightforward and this hierarchical model with such straightforward Gibbs sampling is effective for estimating coefficients and capturing overall patterns in the data. However, it lacks adaptive shrinkage, which means that it treats all coefficients equally and shrinks them all towards zero without allowing variable selection to vary across different clusters (significant and insignificant) of variables.

3.3.2.2 Bayesian Adaptive Lasso

By using the logic of classical Bayesian lasso and adaptive shrinkage idea, we considered the following hierarchical model,

$$\begin{aligned}
\boldsymbol{\theta} \mid \mu_\theta, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\beta}, \kappa &\sim vM(\mu_\theta + g((\boldsymbol{\phi}\boldsymbol{\beta}_{\boldsymbol{\phi}} + \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{X}})), \kappa), \\
\boldsymbol{\beta} \mid \boldsymbol{\tau}^2 &\sim N(0, \mathbf{D}_\tau), \boldsymbol{\tau}^2 = (\tau_1^2, ..., \tau_m^2)\ , \mathbf{D}_\tau = diag(\tau_1^2, ..., \tau_m^2), \\
\kappa &\sim Gamma(a, b), \\
\tau_j^2 &\sim Exp(\lambda_j^2/2)\ , j = 1, ..., m, \\
\lambda_j^2 &\sim Gamma(c, d), \\
\mu_\theta &\sim Circ - Unif(0, 2\pi),
\end{aligned} \tag{3.27}$$

where μ_θ denotes the mean direction of θ, $\boldsymbol{\beta}$ is vector of regression coefficients, κ is concentration parameter. Based on this hierarchical model, full conditional distributions were investigated to implement a Gibbs sampler and given in Appendix A.2.2. This hierarchical model with adaptive shrinkage allows some coefficients to receive heavy shrinkage, effectively being treated as negligible, while allowing other coefficients to remain largely unaffected. This adaptivity is controlled by different shrinkage parameters for each coefficient.

By employing these adaptive shrinkage priors, this hierarchical model can effectively identify and estimate relevant variables by providing better flexibility in capturing

the underlying associations with the response. The Gibbs sampler, combined with these advanced priors, will iteratively sample from the full conditional distributions, incorporating the adaptive shrinkage into the estimation process. This procedure is called as Bayesian Adaptive Lasso, BaLasso in short.

3.3.2.3 Dirichlet Lasso 1

In this section, we treat the variable selection problem as a clustering problem by using the clustering property of Dirichlet Process (DP) priors. This means that we use the DP prior to group the coefficients of independent variables into different clusters based on their similarities as significant or insignificant. This approach can help identify significant variables and group together those with similar effects on the outcome since the DP has the advantage of simultaneously clustering the coefficients of the independent variables. For each tuning parameter $\lambda_j, j = 1, ..., m$, we assume a DP prior distribution since it allows for uncertainty in the number of non-zero coefficients and automatically performs variable selection. We then develop a new version of the lasso, called the Dirichlet lasso 1 (abbreviated as $DLasso_1$), where adaptive shrinkage prior distributions are used for penalizing different coefficients.

By combining the clustering property of DP priors with the adaptive shrinkage of the $DLasso_1$, we aim to obtain a more flexible and effective variable selection approach that can capture complex patterns and relationships among the variables in the data set.

The $DLasso_1$ method is completed by the specification of the following prior distributions:

$$
\begin{aligned}
\boldsymbol{\theta} \mid \mu_\theta, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\beta}, \kappa &\sim vM(\mu_\theta + g((\boldsymbol{\phi}\boldsymbol{\beta}_{\boldsymbol{\phi}} + \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{X}})), \kappa), \\
\boldsymbol{\beta} \mid \boldsymbol{\tau}^2 &\sim N(0, \mathbf{D}_\tau), \boldsymbol{\tau}^2 = (\tau_1^2, ..., \tau_m^2)\ , \mathbf{D}_\tau = diag(\tau_1^2, ..., \tau_m^2), \\
\kappa &\sim Gamma(a, b), \\
\tau_j^2 &\sim Exp(\lambda_j^2/2)\ , j = 1, ..., m, \\
\lambda_j^2 &\sim G, G \sim DP(\alpha, G_0), \\
\mu_\theta &\sim Circ - Unif(0, 2\pi).
\end{aligned}
\tag{3.28}
$$

DP is Dirichlet process prior with the mass parameter $\alpha > 0$ and baseline measure G_0 (Ferguson, 1973). The mass parameter α is also assigned a gamma prior distribution. G_0 can be defined as usual prior distribution such as gamma or half-Cauchy prior. Based on this hierarchical model, full conditional distributions are investigated and provided in Appendix A.2.3.

The $DLasso_1$ is a modification of the classical lasso regression, where adaptive shrinkage prior distributions are used to penalize different coefficients. In classical lasso, the L_1-norm penalty is applied uniformly to all coefficients, which encourages variable selection. In $DLasso_1$, we use adaptive shrinkage priors, likely leveraging the Dirichlet Process, to impose varying levels of shrinkage on different coefficients. This adaptive shrinkage helps in capturing the underlying heterogeneity in the effects of different variables.

3.3.2.4 Dirichlet Lasso 2

In Dirichlet lasso 1 methodology, a DP prior distribution is assumed for each tuning parameter $\lambda_j, j = 1, ..., m$. Now, we propose a method called Dirichlet lasso 2 (abbreviated as $DLasso_2$) which has a new property assuming a mixed prior distribution containing a zero-inflated and normal distribution for each regression coefficient. This proposed method and the one proposed in previous section differ in the following respect:

We assume that regression coefficients are distributed according to conditionally independent identical zero-inflated prior distributions. As a consequence, some of co-

efficient parameters are capable of having large posterior probabilities of being equal to 0 rather than simply being shrunk to 0.

The $DLasso_2$ method is completed by the specification of the following prior distributions:

$$
\begin{aligned}
\boldsymbol{\theta} \mid \mu_\theta, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\beta}, \kappa &\sim vM(\mu_\theta + g((\boldsymbol{\phi}\boldsymbol{\beta}_{\boldsymbol{\phi}} + \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{X}})), \kappa), \\
\boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \boldsymbol{W} &\sim (1 - \boldsymbol{W}) f(*) + (\boldsymbol{W}) N(0, \mathbf{D}_\tau), \\
\boldsymbol{\tau}^2 &= (\tau_1^2, ..., \tau_m^2), \ \mathbf{D}_\tau = diag(\tau_1^2, ..., \tau_m^2), \\
W_j &\sim Bernoulli(\pi_j), \pi_j \sim Beta(\gamma_1, \gamma_2) \ , j = 1, ..., m, \\
\tau_j^2 &\sim Exp(\lambda_j^2/2), \\
\lambda_j^2 &\sim G, G \sim DP(\alpha, G_0), \\
\kappa &\sim Gamma(a, b), \\
\mu_\theta &\sim Circ - Unif(0, 2\pi).
\end{aligned}
\tag{3.29}
$$

where $f(*)$ denotes the point mass at 0, $\boldsymbol{W} = (W_1, W_2, ..., W_m)$ denotes a $m \times 1$ vector of binary random variables with $P(W_j = 1) = \pi_j$. Based on this hierarchical model, full conditional distributions are investigated. The derivations of the full conditional distributions are given in Appendix A.2.4.

In this method, each coefficient is assumed to come from one of the two different distributions, depending on a variable denoted as $\boldsymbol{W}$. The variable $\boldsymbol{W}$ determines the source distribution for each coefficient. If $W_j = 0$, the j^{th} coefficient comes from a distribution around zero, and if $W_j = 1$, it comes from a normal distribution. Again, different shrinkage parameters are considered for each regression coefficient. This allows for adaptive shrinkage, where different coefficients can have different levels of regularization, depending on the source distribution.

3.3.2.5 Dirichlet Mixture Lasso

In Dirichlet lasso 2, a DP prior distribution is assumed for each tuning parameter and also, a mixed prior distribution containing a zero-inflated and normal distribution is

assumed for regression coefficients. In fact, although this prior distribution seems to be a mixed distribution, depending on the value of the W variable, the regression coefficients come from either a distribution around zero or a normal distribution. Now, we relax this assumption of mixed prior distribution by proposing Dirichlet mixture lasso methodology. We assign a mixed of normal densities with different shrinkage parameters for each regression coefficient instead of assigning a mixed prior distribution containing a zero-inflated and normal distribution. In other words, the Dirichlet mixture lasso methodology is a Bayesian approach that extends the Dirichlet lasso 2 method by relaxing the assumption of a mixed prior distribution for regression coefficients. Instead, it assumes that the regression coefficients for the covariates can come from two different distributions, each with its own shrinkage parameter. Here instead of assuming that all regression coefficient follows a normal distribution with different shrinkage parameter, one assumes that regression coefficients for covariates actually can come from two different distributions with different shrinkage parameter, λ_1 and λ_2, where $\lambda_1 \neq \lambda_2$. Moreover, to achieve adaptive shrinkage in this method, we consider different shrinkage parameters for each regression coefficient. Based on this idea, the Dirichlet mixture lasso method is given by the following hierarchical representation:

$$
\begin{aligned}
\boldsymbol{\theta} \mid \mu_\theta, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\beta}, \kappa &\sim vM(\mu_\theta + g((\boldsymbol{\phi}\boldsymbol{\beta}_{\boldsymbol{\phi}} + \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{X}})), \kappa), \\
\boldsymbol{\beta} \mid \boldsymbol{\tau}_1^2, \boldsymbol{\tau}_2^2, \boldsymbol{\pi} &\sim (\boldsymbol{\pi})N(0, \boldsymbol{D}_{\boldsymbol{\tau}_1}) + (1-\boldsymbol{\pi})N(0, \boldsymbol{D}_{\boldsymbol{\tau}_2}), \\
\boldsymbol{\tau}_1^2 &= (\tau_{11}^2, ..., \tau_{1m}^2),\ \boldsymbol{\tau}_2^2 = (\tau_{21}^2, ..., \tau_{2m}^2), \\
\boldsymbol{D}_{\boldsymbol{\tau}_1} &= diag(\tau_{11}^2, \tau_{12}^2, ..., \tau_{1m}^2),\ \boldsymbol{D}_{\boldsymbol{\tau}_2} = diag(\tau_{21}^2, \tau_{22}^2 ..., \tau_{2m}^2), \\
\tau_{1j}^2 &\sim Exp(\lambda_{1j}^2/2),\ \tau_{2j}^2 \sim Exp(\lambda_{2j}^2/2),\ j = 1, ..., m, \\
\lambda_{1j}^2 &\sim G_1, G_1 \sim DP(\alpha, G_{01}), \\
\lambda_{2j}^2 &\sim G_2, G_2 \sim DP(\alpha, G_{02}), \\
\pi_j &\sim Beta(\gamma_1, \gamma_2) \\
\kappa &\sim Gamma(a, b), \\
\mu_\theta &\sim Circ - Unif(0, 2\pi).
\end{aligned}
\tag{3.30}
$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_m)$ denotes a $m \times 1$ vector of mixing probabilities, G_{01} and

G_{02} are base measures and can be defined as gamma or half-Cauchy distribution for robust inference. Note that mixing probabilities π_j can be different for each coefficient. Based on this hierarchical model, full conditional distributions are investigated. The derivations of the full conditional distributions are provided in Appendix A.2.5.

The advantage of the Dirichlet mixture lasso methodology lies in its flexibility and ability to capture different patterns of sparsity and shrinkage in the regression coefficients. By allowing each coefficient to come from one or both of two distributions with different shrinkage parameters, it can better adapt to complex data structures and improve predictive performance.

Overall, the Dirichlet mixture lasso methodology offers an alternative Bayesian approach to variable selection and regularization in regression models, providing a more flexible and adaptive way to handle the sparsity and shrinkage of regression coefficients.

CHAPTER 4

NUMERICAL STUDIES

In this chapter, we conduct extensive simulation studies showing the performances of the proposed variable selection methodologies under different realistic scenarios and illustrate their applications in real datasets. By simulation studies and analyzing real datasets, the chapter aims to demonstrate the effectiveness and usefulness of the proposed methodologies for various practical applications and data scenarios.

4.1 Simulation Study

In this section, we provide some simulation studies to examine the performances and behavior of the proposed methodologies first in the context of linear-circular regression models and then for circular-(circular, linear) regression models. All this is accomplished by orchestrating a comprehensive Monte Carlo (MC) study. In simulation studies, we use quite common technicality in the realm of Bayesian simulation studies, where data generation process is carried out in R (R Core Team, 2023) and MCMC method is implemented using MultiBUGS (Goudie et al., 2020) which is an open source software for Bayesian statistics. R programming language and MultiBUGS program are integrated to carry out all the analyses. This integration allows us to perform sophisticated Bayesian simulation studies efficiently.

Here's an overview of how this integration typically works:

- Data Generation in R: In the simulation studies, we create data in R. This could involve generating data from known probability distributions or using specific models with known parameter set that mimic real-world scenarios.

- Model Specification in MultiBUGS: Once we have the data simulated in R, we need to define the Bayesian model in MultiBUGS. This involves specifying the likelihood function and prior distributions for all the parameters of interest. MultiBUGS uses the BUGS language, which is graphical based a probabilistic programming language designed specifically for Bayesian modeling.

- MCMC Sampling in MultiBUGS: After specifying the model, MCMC methods, such as the Metropolis-Hastings algorithm or Gibbs sampling and their versions, are used to draw samples from the posterior distribution of the parameters given the sampling distribution and the prior distributions.

- Interface with R: The R programming language is used to interact with MultiBUGS. We use R to prepare the data, pass the data to MultiBUGS, and execute the MCMC sampling. MultiBUGS then runs the MCMC chains and returns the posterior samples back to R.

- Posterior Analysis in R: Once we have the posterior samples from MultiBUGS, we can perform various analyses in R to summarize and interpret the results. This may involve calculating posterior means, credible intervals, and other quantities of interest.

- Visualization and Reporting: R provides various visualization tools to plot the results and communicate the findings effectively.

In all simulation studies, we consider two different chains with 90,000 updates after discarding the first 10,000 iterations as burn-in (warm-up) period. These two different MCMC chains are run in parallel. Each chain performs a sequence of updates to explore the posterior distribution of the model parameters. Trace plots and Brooks-Gelman-Rubin (BGR) statistics (Brooks and Gelman, 1998) are used to monitor convergence of the chains as well as to determine warm-up period and number of MCMC samples to be used for final posterior inference. The chains are run until the Monte Carlo errors based on the Markov chain become less than 5% of the posterior standard deviations. This is an indication that the chains have adequately approximated the posterior distributions. Posterior means (i.e. expectation of the posterior distributions) and 95% equal-tailed posterior credible intervals are used for estimating

the model parameters and for variable selection for symmetric posterior distributions. For asymmetric posterior distributions, posterior medians and interquartile ranges are used. Variables whose credible intervals do not include zero are considered to be statistically significant and is included in the final model.

4.1.1 Performance Metrics

Performance Metrics for Variable Selection: Various performance metrics are used to evaluate the performances of the variable selection methodologies investigated here. Specific definitions and calculations of these measures are as follow:

- **Average Correct Inclusion (ACI):** This metric measures the method's performance of correctly including the covariates in the model. It estimates the expected number of covariates correctly included in the model by a given method. A covariate is considered correctly included if it was used in the true response data generating mechanism and is also found to be significant in the model. It is given by

$$ACI = \frac{\sum_{i=1}^{M} CI_i}{M} \tag{4.1}$$

  where M is the number of MC replications and CI_i is the number of covariates that are correctly included in the model in i^{th} MC replication.

  ACI provides a useful measure to assess how well a variable selection method performs in capturing the important covariates.

- **Average False Exclusion (AFE):** Average false exclusion is another performance metric used to evaluate the effectiveness of a variable selection method. It measures the method in terms of incorrectly excluding the covariates in the model. It estimates the expected number of covariates incorrectly excluded from the model by a given method. Incorrect exclusion refers to the method excluding the covariates that were actually a part of the true response data generating mechanism. It is given by

$$AFE = \frac{\sum_{i=1}^{M} FE_i}{M} \tag{4.2}$$

where FE_i is the number of covariates that are falsely excluded from the model in i^{th} MC replication.

By calculating both ACI and AFE, we can gain a understanding of the strengths and weaknesses of the variable selection method. A good method should have a high ACI and a low AFE, indicating that it correctly includes important covariates and avoids excluding significant predictors.

- **Correct Inclusion Ratio (CIR)):** Correct inclusion ratio is another performance metric used in the evaluation of variable selection methods. Ratio of replications where all covariates are correctly included in the model to all replications and is given by

$$CIR = \frac{\sum_{i=1}^{M} CII_i}{M} \tag{4.3}$$

where CII_i is the indicator of correct inclusion for i^{th} MC replication. CII can take two different values for each replication. CII_i is 1 if all covariates are correctly included, 0 otherwise.

A perfect method would have a CIR of 1, indicating that all the relevant covariates were correctly included in the model. A lower CIR indicates that the method failed to include some important covariates.

Performance Metrics for Parameter Estimation: To evaluate the performance of parameter estimation, two commonly used performance metrics are Relative Bias (RB) and Mean Squared Error (MSE) and are given below.

- **Relative Bias:** Relative Bias measures bias adjusted for the magnitude of the true parameter value and is unit-free. Because of its unitless nature, it enables comparison across covariates. It is defined as follows:

$$RB = \frac{E(\hat{\beta}) - \beta}{\beta}, \tag{4.4}$$

where, β is the true value and $\hat{\beta}$ is an estimate.

- **Mean Squared Error:** MSE gives a measure for expected loss due to the estimator used and is given by $MSE = E[(\hat{\beta} - \beta)^2]$.

  MSE is a measure of the overall accuracy and precision of the estimator.

4.1.2 Simulation Study: Bayesian Lasso in Linear-Circular Regression Models

This section first investigates the sensitivity of the posterior results to standard λ hyper-priors (Section 4.1.2.2) and the novel EB-GS hyper-prior (Section 4.1.2.3). Then, it investigates the performance of the proposed EB-GS method in parameter estimation and variable selection (Section 4.1.2.4).

4.1.2.1 Simulation Design

Our simulation scenarios mimic those in the original lasso paper (Tibshirani, 1996). Our design is controlled for the sample size and the heterogeneity of the circular covariate data. In all simulation settings, circular covariates are independently generated from the von Mises distributions with circular mean 0 without loss of generality and various different concentration parameters, $\kappa = 2, 4, 6, 8, 10$. Some circular distributions with small κ may require special statistical treatment in the analysis, therefore, we also considered smaller values of concentration parameter ($\kappa = 0.5, 1, 1.5, 1.75, 1.99$) in the simulation studies that investigates the performance of the proposed EB-GS method. Linear response variables, Y_i's, are generated from the model given in Section 3.3.1 with $\epsilon_i \sim N(0, 9)$. The study is controlled for the sample size using, n=100 and 250.

The three simulation scenarios considered are as follows:

- **Scenario I**: $\beta = (4, 3, 3, 1.5, 1.5, 0, 0, 2, 2, 0, 0)$, p=5.
- **Scenario II**: $\beta = (4, 0.85, 0.85, 0.85, 0.85, 0, 0, 0.85, 0.850, 0)$, p=5.
- **Scenario III**: $\beta = (4, \underbrace{3, ..., 3}_{16}, \underbrace{0, ..., 0}_{16})$, p=16.

where p is the number of circular covariates. In these scenarios, the value of β_0 is set to 4. We considered a variety of hyper-prior distributions for tuning parameter λ including non-informative, weakly informative and informative priors. Priors for the intercept β_0 and error variance σ^2 are specified as $\beta_0 \sim N(0, 100)$ and $\sigma^2 \sim Ga(0.001, 0.001)$. Each scenario is repeated 500 times.

4.1.2.2 Sensitivity analysis for classical Bayesian lasso

We conducted a detailed sensitivity analysis to assess if and how the performance of the posterior estimates in linear-circular regression models are affected by conventional hyper prior settings for λ. One such prior is half-Cauchy hyper-prior which is becoming increasingly popular in the Bayesian literature as a robust alternative (Gelman, 2006; Polson and Scott, 2012; Mulder and Pericchi, 2018). Another one is gamma hyper-prior which is proposed by Park and Casella (2008) as a suitable hyper-prior for the tuning parameter. Table 4.1 lists the hyper-priors considered in this study taking the account for different degrees of informativeness.

Table 4.1: Conventional Hyper-prior distributions for the tuning parameter.

1. Ga(1,0.1)
2. Ga(0.1, 0.1)
3. Ga(0.01, 0.01)
4. Ga(0.001, 0.001)
5. HC(0, 1)
6. HC(0, 1.2)
7. HC(0, 1.5)
8. HC(0, 1.7)
9. HC(0, 2)

Tables 4.2-4.5 give RB and MSE of coefficient estimates under various different hyper-prior setting for λ. MSEs are presented in the parentheses in all tables. Accordingly, in general, better coefficient estimation (low RB and MSE) are associated with more informative λ priors when gamma type prior distributions are used. When

HC type priors are used, RBs and MSEs are comparable. Comparing gamma and HC type priors, it is seen that the statistical properties of the final estimates are similar. Overall, the simulation study shows that prior setting for λ effects final parameter estimation in linear-circular regression particularly when the tuning parameter has a gamma type hyper-prior.

Performance measures for variable selection are given in Tables 4.6-4.8. First of all, the true behavior for variable selection is as follows. For scenarios I and II, ACI should be 3, and for scenario III it should be 8. AFE and CIR should be 0 and 1 respectively in all scenarios. In general, more informative hyper-priors seem to lead to a clearer distinction between significant and insignificant covariates rendering more qualified variable selection. However this is not necessarily the case with less informative hyper-priors. Overall, performance of Bayesian lasso in linear-circular regression is sensitive to hyper-prior choice for λ.

Table 4.2: Relative Bias (MSE) for Scenario I.

n	Coefficients	Ga(0.001, 0.001)	Ga(0.01, 0.01)	Ga(0.1, 0.1)	Ga(1, 0.1)	HC(0, 1)	HC(0, 1.2)	HC(0, 1.5)	HC(0, 1.7)	HC(0, 2)
100	β_0	0.36(5.49)	0.35(5.45)	0.31(5.23)	0.42(5.54)	0.33(5.35)	0.33(5.35)	0.35(5.32)	0.35(5.36)	0.37(5.42)
	β_1	-0.13(0.87)	-0.13(0.86)	-0.12(0.81)	-0.15(0.89)	-0.12(0.83)	-0.13(0.84)	-0.13(0.85)	-0.13(0.85)	-0.13(0.86)
	β_2	-0.07(0.33)	-0.07(0.33)	-0.06(0.31)	-0.08(0.34)	-0.07(0.32)	-0.07(0.32)	-0.07(0.32)	-0.07(0.33)	-0.07(0.33)
	β_3	-0.37(1.2)	-0.36(1.2)	-0.33(1.22)	-0.41(1.15)	-0.34(1.22)	-0.34(1.21)	-0.36(1.19)	-0.36(1.20)	-0.37(1.19)
	β_4	-0.19(0.44)	-0.18(0.43)	-0.17(0.42)	-0.20(0.44)	-0.17(0.43)	-0.17(0.43)	-0.18(0.43)	-0.18(0.43)	-0.18(0.43)
	β_5	NA(1.01)	NA(1.03)	NA(1.14)	NA(0.76)	NA(1.12)	NA(1.08)	NA(1.02)	NA(1.00)	NA(0.97)
	β_6	NA(0.32)	NA(0.32)	NA(0.34)	NA(0.29)	NA(0.33)	NA(0.33)	NA(0.32)	NA(0.32)	NA(0.32)
	β_7	-0.55(2.49)	-0.54(2.50)	-0.51(2.54)	-0.62(2.42)	-0.52(2.54)	-0.53(2.52)	-0.54(2.49)	-0.55(2.48)	-0.56(2.47)
	β_8	-0.25(0.94)	-0.24(0.93)	-0.22(0.9)	-0.27(0.95)	-0.23(0.91)	-0.23(0.92)	-0.24(0.92)	-0.24(0.93)	-0.25(0.93)
	β_9	NA(0.99)	NA(1.00)	NA(1.15)	NA(0.63)	NA(1.13)	NA(1.07)	NA(1.00)	NA(0.96)	NA(0.92)
	β_{10}	NA(0.54)	NA(0.55)	NA(0.59)	NA(0.49)	NA(0.58)	NA(0.57)	NA(0.56)	NA(0.55)	NA(0.54)
250	β_0	0.27(3.96)	0.27(3.97)	0.24(3.92)	0.31(3.90)	0.25(4.04)	0.26(3.95)	0.26(3.92)	0.27(3.92)	0.27(3.94)
	β_1	-0.05(0.30)	-0.05(0.29)	-0.05(0.29)	-0.06(0.30)	-0.05(0.29)	-0.05(0.29)	-0.05(0.29)	-0.05(0.29)	-0.05(0.29)
	β_2	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)
	β_3	-0.22(0.75)	-0.22(0.75)	-0.20(0.75)	-0.25(0.74)	-0.21(0.75)	-0.21(0.75)	-0.21(0.75)	-0.22(0.75)	-0.22(0.75)
	β_4	-0.09(0.19)	-0.09(0.19)	-0.08(0.19)	-0.09(0.20)	-0.08(0.19)	-0.08(0.19)	-0.08(0.19)	-0.08(0.19)	-0.09(0.19)
	β_5	NA(0.63)	NA(0.63)	NA(0.68)	NA(0.54)	NA(0.72)	NA(0.66)	NA(0.64)	NA(0.63)	NA(0.62)
	β_6	NA(0.17)	NA(0.17)	NA(0.18)	NA(0.16)	NA(0.18)	NA(0.17)	NA(0.17)	NA(0.17)	NA(0.17)
	β_7	-0.46(2.25)	-0.46(2.25)	-0.43(2.26)	-0.51(2.21)	-0.44(2.28)	-0.44(2.26)	-0.45(2.25)	-0.45(2.24)	-0.46(2.25)
	β_8	-0.10(0.37)	-0.10(0.37)	-0.09(0.36)	-0.11(0.38)	-0.09(0.36)	-0.10(0.36)	-0.10(0.37)	-0.10(0.37)	-0.10(0.37)
	β_9	NA(0.92)	NA(0.92)	NA(1.04)	NA(0.70)	NA(1.02)	NA(0.97)	NA(0.93)	NA(0.92)	NA(0.89)
	β_{10}	NA(0.25)	NA(0.25)	NA(0.26)	NA(0.23)	NA(0.26)	NA(0.25)	NA(0.25)	NA(0.25)	NA(0.25)

NA: No results are available for these parameters.

Table 4.3: Relative Bias (MSE) for Scenario II.

n	Coefficients	Ga(0.001, 0.001)	Ga(0.01, 0.01)	Ga(0.1, 0.1)	Ga(1, 0.1)	HC(0, 1)	HC(0, 1.2)	HC(0, 1.5)	HC(0, 1.7)	HC(0, 2)
100	β_0	0.34(2.65)	0.31(2.44)	0.23(2.43)	0.28(2.26)	0.29(2.50)	0.30(2.50)	0.31(2.50)	0.31(2.50)	0.31(2.51)
	β_1	-0.48(0.39)	-0.40(0.37)	-0.28(0.38)	-0.35(0.36)	-0.40(0.38)	-0.41(0.38)	-0.42(0.38)	-0.42(0.38)	-0.43(0.38)
	β_2	-0.23(0.27)	-0.14(0.25)	-0.02(0.25)	-0.07(0.24)	-0.15(0.27)	-0.15(0.26)	-0.16(0.27)	-0.17(0.26)	-0.18(0.27)
	β_3	-0.64(0.65)	-0.58(0.66)	-0.44(0.80)	-0.54(0.66)	-0.56(0.70)	-0.57(0.70)	-0.58(0.68)	-0.58(0.68)	-0.59(0.67)
	β_4	-0.37(0.33)	-0.28(0.31)	-0.15(0.32)	-0.22(0.30)	-0.29(0.32)	-0.29(0.32)	-0.30(0.32)	-0.31(0.32)	-0.32(0.32)
	β_5	NA(0.22)	NA(0.29)	NA(0.60)	NA(0.32)	NA(0.36)	NA(0.34)	NA(0.32)	NA(0.30)	NA(0.28)
	β_6	NA(0.15)	NA(0.20)	NA(0.31)	NA(0.25)	NA(0.21)	NA(0.20)	NA(0.20)	NA(0.19)	NA(0.19)
	β_7	-0.85(0.68)	-0.82(0.69)	-0.73(0.83)	-0.81(0.68)	-0.8(0.73)	-0.81(0.72)	-0.82(0.71)	-0.82(0.69)	-0.83(0.69)
	β_8	-0.51(0.43)	-0.42(0.41)	-0.28(0.43)	-0.36(0.40)	-0.42(0.42)	-0.43(0.42)	-0.44(0.42)	-0.44(0.42)	-0.45(0.42)
	β_9	NA(0.11)	NA(0.15)	NA(0.37)	NA(0.17)	NA(0.20)	NA(0.19)	NA(0.17)	NA(0.16)	NA(0.15)
	β_{10}	NA(0.18)	NA(0.23)	NA(0.36)	NA(0.28)	NA(0.25)	NA(0.25)	NA(0.23)	NA(0.23)	NA(0.22)
250	β_0	0.30(1.89)	0.27(1.76)	0.2(1.78)	0.24(1.65)	0.27(1.80)	0.27(1.81)	0.27(1.77)	0.28(1.81)	0.28(1.80)
	β_1	-0.31(0.22)	-0.26(0.21)	-0.17(0.20)	-0.21(0.20)	-0.26(0.21)	-0.27(0.22)	-0.27(0.22)	-0.27(0.22)	-0.28(0.22)
	β_2	-0.17(0.12)	-0.12(0.10)	-0.05(0.09)	-0.08(0.10)	-0.13(0.11)	-0.13(0.11)	-0.13(0.11)	-0.14(0.11)	-0.14(0.11)
	β_3	-0.58(0.45)	-0.53(0.43)	-0.40(0.43)	-0.47(0.42)	-0.52(0.44)	-0.52(0.44)	-0.52(0.44)	-0.53(0.44)	-0.53(0.44)
	β_4	-0.22(0.19)	-0.17(0.18)	-0.08(0.17)	-0.11(0.17)	-0.17(0.18)	-0.17(0.19)	-0.17(0.19)	-0.18(0.19)	-0.18(0.19)
	β_5	NA(0.15)	NA(0.20)	NA(0.41)	NA(0.27)	NA(0.23)	NA(0.22)	NA(0.21)	NA(0.20)	NA(0.19)
	β_6	NA(0.10)	NA(0.11)	NA(0.14)	NA(0.13)	NA(0.11)	NA(0.11)	NA(0.11)	NA(0.11)	NA(0.11)
	β_7	-0.90(0.77)	-0.87(0.80)	-0.78(0.97)	-0.83(0.81)	-0.87(0.85)	-0.87(0.84)	-0.87(0.82)	-0.88(0.82)	-0.88(0.80)
	β_8	-0.49(0.35)	-0.45(0.33)	-0.36(0.33)	-0.40(0.32)	-0.44(0.34)	-0.44(0.34)	-0.45(0.34)	-0.45(0.34)	-0.46(0.34)
	β_9	NA(0.09)	NA(0.12)	NA(0.30)	NA(0.17)	NA(0.14)	NA(0.14)	NA(0.13)	NA(0.12)	NA(0.12)
	β_{10}	NA(0.16)	NA(0.19)	NA(0.26)	NA(0.22)	NA(0.20)	NA(0.19)	NA(0.19)	NA(0.19)	NA(0.18)

NA: No results are available for these parameters.

Table 4.4: Relative Bias (MSE) for Scenario III, n=100.

Coefficients	Ga(0.001, 0.001)	Ga(0.01, 0.01)	Ga(0.1, 0.1)	Ga(1, 0.1)	HC(0, 1)	HC(0, 1.2)	HC(0, 1.5)	HC(0, 1.7)	HC(0, 2)
β_0	1.71(58.55)	1.7(58)	1.69(57.32)	1.83(64.13)	1.67(56.50)	1.69(56.96)	1.71(58.18)	1.72(58.54)	1.73(58.84)
β_1	-0.13(0.92)	-0.13(0.92)	-0.13(0.91)	-0.14(0.93)	-0.13(0.91)	-0.13(0.91)	-0.13(0.91)	-0.13(0.91)	-0.13(0.92)
β_2	-0.08(0.44)	-0.08(0.44)	-0.08(0.44)	-0.08(0.45)	-0.08(0.44)	-0.08(0.44)	-0.08(0.44)	-0.08(0.44)	-0.08(0.44)
β_3	-0.39(2.88)	-0.39(2.88)	-0.38(2.87)	-0.41(2.95)	-0.38(2.86)	-0.38(2.86)	-0.39(2.88)	-0.39(2.88)	-0.39(2.88)
β_4	-0.09(0.67)	-0.09(0.67)	-0.09(0.66)	-0.10(0.68)	-0.09(0.66)	-0.09(0.66)	-0.09(0.66)	-0.09(0.67)	-0.09(0.67)
β_5	-0.48(4.08)	-0.48(4.07)	-0.48(4.06)	-0.51(4.11)	-0.47(4.08)	-0.48(4.08)	-0.48(4.08)	-0.48(4.09)	-0.49(4.09)
β_6	-0.17(1.18)	-0.17(1.18)	-0.16(1.18)	-0.18(1.21)	-0.16(1.17)	-0.16(1.17)	-0.16(1.18)	-0.17(1.18)	-0.17(1.18)
β_7	-0.58(4.75)	-0.58(4.76)	-0.57(4.76)	-0.60(4.78)	-0.57(4.75)	-0.57(4.75)	-0.57(4.76)	-0.58(4.75)	-0.58(4.76)
β_8	-0.21(1.43)	-0.21(1.43)	-0.21(1.42)	-0.23(1.47)	-0.21(1.41)	-0.21(1.41)	-0.21(1.42)	-0.21(1.42)	-0.21(1.43)
β_9	-0.13(0.88)	-0.13(0.88)	-0.13(0.88)	-0.14(0.90)	-0.12(0.87)	-0.13(0.87)	-0.13(0.88)	-0.13(0.88)	-0.13(0.88)
β_{10}	-0.08(0.37)	-0.08(0.37)	-0.08(0.37)	-0.09(0.37)	-0.08(0.36)	-0.08(0.36)	-0.08(0.37)	-0.08(0.37)	-0.08(0.37)
β_{11}	-0.35(2.73)	-0.34(2.72)	-0.34(2.72)	-0.37(2.77)	-0.34(2.71)	-0.34(2.71)	-0.34(2.72)	-0.35(2.73)	-0.35(2.73)
β_{12}	-0.11(0.69)	-0.11(0.69)	-0.11(0.68)	-0.12(0.70)	-0.11(0.68)	-0.11(0.68)	-0.11(0.68)	-0.11(0.68)	-0.11(0.69)
β_{13}	-0.52(4.16)	-0.52(4.16)	-0.52(4.15)	-0.54(4.20)	-0.51(4.14)	-0.51(4.14)	-0.52(4.16)	-0.52(4.17)	-0.52(4.15)
β_{14}	-0.13(0.97)	-0.12(0.96)	-0.12(0.96)	-0.14(0.98)	-0.12(0.95)	-0.12(0.96)	-0.12(0.96)	-0.12(0.96)	-0.13(0.96)
β_{15}	-0.60(4.66)	-0.60(4.64)	-0.60(4.63)	-0.63(4.74)	-0.60(4.63)	-0.60(4.63)	-0.60(4.65)	-0.60(4.66)	-0.61(4.65)
β_{16}	-0.18(1.19)	-0.18(1.18)	-0.18(1.18)	-0.19(1.22)	-0.18(1.17)	-0.18(1.17)	-0.18(1.18)	-0.18(1.18)	-0.18(1.19)
β_{17}	NA(0.29)	NA(0.29)	NA(0.30)	NA(0.28)	NA(0.30)	NA(0.30)	NA(0.30)	NA(0.29)	NA(0.29)
β_{18}	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.23)	NA(0.23)	NA(0.22)	NA(0.22)	NA(0.22)
β_{19}	NA(0.89)	NA(0.89)	NA(0.90)	NA(0.80)	NA(0.92)	NA(0.91)	NA(0.90)	NA(0.89)	NA(0.88)
β_{20}	NA(0.36)	NA(0.37)	NA(0.37)	NA(0.35)	NA(0.37)	NA(0.37)	NA(0.37)	NA(0.36)	NA(0.36)
β_{21}	NA(0.96)	NA(0.98)	NA(0.99)	NA(0.85)	NA(1.01)	NA(1.00)	NA(0.97)	NA(0.97)	NA(0.95)
β_{22}	NA(0.41)	NA(0.41)	NA(0.41)	NA(0.39)	NA(0.41)	NA(0.41)	NA(0.41)	NA(0.41)	NA(0.41)
β_{23}	NA(1.20)	NA(1.20)	NA(1.23)	NA(1.04)	NA(1.27)	NA(1.25)	NA(1.21)	NA(1.21)	NA(1.18)
β_{24}	NA(0.53)	NA(0.54)	NA(0.54)	NA(0.51)	NA(0.55)	NA(0.54)	NA(0.54)	NA(0.54)	NA(0.53)
β_{25}	NA(0.40)	NA(0.40)	NA(0.40)	NA(0.39)	NA(0.41)	NA(0.41)	NA(0.40)	NA(0.40)	NA(0.40)
β_{26}	NA(0.21)	NA(0.21)	NA(0.21)	NA(0.20)	NA(0.21)	NA(0.21)	NA(0.21)	NA(0.21)	NA(0.21)
β_{27}	NA(0.93)	NA(0.94)	NA(0.94)	NA(0.86)	NA(0.96)	NA(0.95)	NA(0.94)	NA(0.93)	NA(0.92)
β_{28}	NA(0.34)	NA(0.34)	NA(0.34)	NA(0.32)	NA(0.34)	NA(0.34)	NA(0.34)	NA(0.34)	NA(0.34)
β_{29}	NA(1.19)	NA(1.20)	NA(1.22)	NA(1.05)	NA(1.25)	NA(1.23)	NA(1.21)	NA(1.20)	NA(1.18)
β_{30}	NA(0.38)	NA(0.38)	NA(0.39)	NA(0.37)	NA(0.39)	NA(0.39)	NA(0.39)	NA(0.39)	NA(0.38)
β_{31}	NA(1.06)	NA(1.07)	NA(1.08)	NA(0.91)	NA(1.11)	NA(1.10)	NA(1.06)	NA(1.05)	NA(1.05)
β_{32}	NA(0.55)	NA(0.55)	NA(0.56)	NA(0.52)	NA(0.56)	NA(0.56)	NA(0.55)	NA(0.55)	NA(0.55)

NA: No results are available for these parameters.

Table 4.5: Relative Bias (MSE) for Scenario III, n=250.

Coefficients	Ga(0.001, 0.001)	Ga(0.01, 0.01)	Ga(0.1, 0.1)	Ga(1, 0.1)	HC(0, 1)	HC(0, 1.2)	HC(0, 1.5)	HC(0, 1.7)	HC(0, 2)
β_0	1.29(35.31)	1.31(35.87)	1.27(34.22)	1.35(38.16)	1.27(34.46)	1.29(35.2)	1.29(35.18)	1.31(35.73)	1.3(35.55)
β_1	-0.04(0.27)	-0.04(0.27)	-0.04(0.26)	-0.04(0.27)	-0.04(0.26)	-0.04(0.27)	-0.04(0.27)	-0.04(0.27)	-0.04(0.27)
β_2	-0.03(0.09)	-0.03(0.09)	-0.03(0.09)	-0.03(0.09)	-0.03(0.09)	-0.03(0.09)	-0.03(0.09)	-0.03(0.09)	-0.03(0.09)
β_3	-0.10(0.93)	-0.10(0.94)	-0.09(0.92)	-0.11(0.95)	-0.09(0.93)	-0.10(0.93)	-0.10(0.93)	-0.10(0.94)	-0.10(0.93)
β_4	-0.05(0.16)	-0.05(0.16)	-0.05(0.16)	-0.05(0.16)	-0.05(0.16)	-0.05(0.16)	-0.05(0.16)	-0.05(0.16)	-0.05(0.16)
β_5	-0.40(2.56)	-0.40(2.58)	-0.40(2.54)	-0.41(2.62)	-0.40(2.52)	-0.40(2.55)	-0.40(2.56)	-0.40(2.57)	-0.40(2.57)
β_6	-0.07(0.23)	-0.07(0.23)	-0.07(0.23)	-0.07(0.23)	-0.07(0.23)	-0.07(0.23)	-0.07(0.23)	-0.07(0.23)	-0.07(0.23)
β_7	-0.47(3.71)	-0.48(3.80)	-0.47(3.76)	-0.49(3.84)	-0.47(3.81)	-0.47(3.75)	-0.47(3.78)	-0.48(3.79)	-0.47(3.80)
β_8	-0.07(0.43)	-0.07(0.43)	-0.06(0.43)	-0.07(0.43)	-0.06(0.42)	-0.06(0.43)	-0.06(0.43)	-0.07(0.43)	-0.07(0.43)
β_9	-0.05(0.21)	-0.05(0.21)	-0.05(0.21)	-0.06(0.21)	-0.05(0.21)	-0.05(0.21)	-0.05(0.21)	-0.05(0.21)	-0.05(0.21)
β_{10}	-0.02(0.11)	-0.02(0.11)	-0.02(0.11)	-0.02(0.11)	-0.02(0.11)	-0.02(0.11)	-0.02(0.11)	-0.02(0.11)	-0.02(0.11)
β_{11}	-0.18(1.59)	-0.19(1.62)	-0.18(1.60)	-0.19(1.62)	-0.18(1.60)	-0.18(1.61)	-0.18(1.61)	-0.18(1.6)	-0.19(1.61)
β_{12}	-0.07(0.25)	-0.07(0.25)	-0.07(0.25)	-0.07(0.25)	-0.07(0.25)	-0.07(0.25)	-0.07(0.25)	-0.07(0.25)	-0.07(0.25)
β_{13}	-0.38(2.39)	-0.38(2.38)	-0.38(2.36)	-0.40(2.47)	-0.38(2.37)	-0.38(2.38)	-0.38(2.38)	-0.39(2.39)	-0.39(2.42)
β_{14}	-0.08(0.27)	-0.08(0.27)	-0.08(0.27)	-0.08(0.27)	-0.08(0.27)	-0.08(0.27)	-0.08(0.27)	-0.08(0.27)	-0.08(0.27)
β_{15}	-0.52(3.63)	-0.53(3.64)	-0.51(3.60)	-0.54(3.71)	-0.51(3.62)	-0.52(3.61)	-0.52(3.64)	-0.52(3.67)	-0.52(3.64)
β_{16}	-0.08(0.34)	-0.08(0.34)	-0.08(0.34)	-0.09(0.34)	-0.08(0.34)	-0.08(0.34)	-0.08(0.34)	-0.08(0.34)	-0.08(0.34)
β_{17}	NA(0.17)	NA(0.17)	NA(0.17)	NA(0.16)	NA(0.17)	NA(0.17)	NA(0.17)	NA(0.17)	NA(0.17)
β_{18}	NA(0.13)	NA(0.13)	NA(0.13)	NA(0.13)	NA(0.13)	NA(0.13)	NA(0.13)	NA(0.13)	NA(0.13)
β_{19}	NA(0.50)	NA(0.50)	NA(0.51)	NA(0.49)	NA(0.51)	NA(0.51)	NA(0.51)	NA(0.51)	NA(0.49)
β_{20}	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)
β_{21}	NA(1.13)	NA(1.12)	NA(1.16)	NA(1.06)	NA(1.14)	NA(1.15)	NA(1.15)	NA(1.12)	NA(1.11)
β_{22}	NA(0.12)	NA(0.12)	NA(0.12)	NA(0.12)	NA(0.12)	NA(0.12)	NA(0.12)	NA(0.12)	NA(0.12)
β_{23}	NA(0.46)	NA(0.46)	NA(0.46)	NA(0.40)	NA(0.47)	NA(0.46)	NA(0.46)	NA(0.45)	NA(0.45)
β_{24}	NA(0.26)	NA(0.26)	NA(0.26)	NA(0.26)	NA(0.26)	NA(0.26)	NA(0.26)	NA(0.26)	NA(0.26)
β_{25}	NA(0.24)	NA(0.24)	NA(0.24)	NA(0.23)	NA(0.24)	NA(0.24)	NA(0.24)	NA(0.24)	NA(0.24)
β_{26}	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)	NA(0.10)
β_{27}	NA(0.54)	NA(0.54)	NA(0.55)	NA(0.52)	NA(0.55)	NA(0.55)	NA(0.54)	NA(0.54)	NA(0.54)
β_{28}	NA(0.09)	NA(0.09)	NA(0.09)	NA(0.09)	NA(0.09)	NA(0.09)	NA(0.09)	NA(0.09)	NA(0.09)
β_{29}	NA(0.88)	NA(0.91)	NA(0.92)	NA(0.82)	NA(0.92)	NA(0.91)	NA(0.9)	NA(0.88)	NA(0.89)
β_{30}	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.22)
β_{31}	NA(0.78)	NA(0.79)	NA(0.82)	NA(0.71)	NA(0.83)	NA(0.83)	NA(0.79)	NA(0.78)	NA(0.77)
β_{32}	NA(0.14)	NA(0.14)	NA(0.14)	NA(0.14)	NA(0.14)	NA(0.14)	NA(0.14)	NA(0.14)	NA(0.14)

NA: No results are available for these parameters.

Table 4.6: Performances in variable selection under Scenario I.

n	Measures	Ga(0.001, 0.001)	Ga(0.01, 0.01)	Ga(0.1, 0.1)	Ga(1, 0.1)	HC(0, 1)	HC(0, 1.2)	HC(0, 1.5)	HC(0, 1.7)	HC(0, 2)
100	ACI	1.832	1.840	1.890	1.806	1.862	1.858	1.850	1.850	1.846
	AFE	1.168	1.160	1.110	1.194	1.138	1.142	1.150	1.150	1.154
	CIR	0.180	0.176	0.196	0.162	0.186	0.182	0.178	0.178	0.178
250	ACI	2.822	2.831	2.840	2.817	2.836	2.840	2.831	2.836	2.836
	AFE	0.178	0.169	0.160	0.183	0.164	0.160	0.169	0.164	0.164
	CIR	0.850	0.850	0.850	0.845	0.850	0.850	0.850	0.850	0.850

True values are ACI=3, AFE=0, CIR=1.

Table 4.7: Performances in variable selection under Scenario II.

n	Measures	Ga(0.001, 0.001)	Ga(0.01, 0.01)	Ga(0.1, 0.1)	Ga(1, 0.1)	HC(0, 1)	HC(0, 1.2)	HC(0, 1.5)	HC(0, 1.7)	HC(0, 2)
100	ACI	0.351	0.471	0.672	0.562	0.455	0.451	0.422	0.422	0.406
	AFE	2.649	2.529	2.328	2.438	2.545	2.549	2.578	2.578	2.594
	CIR	0.006	0.006	0.013	0.003	0.010	0.010	0.010	0.006	0.006
250	ACI	1.082	1.233	1.397	1.356	1.205	1.219	1.219	1.219	1.178
	AFE	1.918	1.767	1.603	1.644	1.795	1.781	1.781	1.781	1.822
	CIR	0.027	0.027	0.041	0.027	0.027	0.027	0.027	0.027	0.027

True values are ACI=3, AFE=0, CIR=1

Table 4.8: Performances in variable selection under Scenario III.

n	Measures	Ga(0.001, 0.001)	Ga(0.01, 0.01)	Ga(0.1, 0.1)	Ga(1, 0.1)	HC(0, 1)	HC(0, 1.2)	HC(0, 1.5)	HC(0, 1.7)	HC(0, 2)
100	ACI	6.924	6.927	6.948	6.863	6.966	6.957	6.951	6.933	6.912
	AFE	1.076	1.073	1.052	1.137	1.034	1.043	1.049	1.067	1.088
	CIR	0.357	0.351	0.366	0.320	0.360	0.360	0.357	0.354	0.341
250	ACI	8.000	8.000	8.000	8.000	8.000	8.000	8.000	8.000	8.000
	AFE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	CIR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

True values are ACI=8, AFE=0, CIR=1

4.1.2.3 Sensitivity analysis for EB-GS hyper-prior for λ

In this part, we investigate sensitivity of the novel EB-GS based Bayesian lasso results to hyper-hyper prior settings assumed for r and δ parameters. This sensitivity analysis is made by considering just Scenario I. We consider that the effects of the coefficients in this scenario will measure the sensitivity of the EB-GS method. The hyper-hyper priors taking the account for different degrees of non-informativeness are as follows: Ga(0.01,0.01) and Ga(0.1,0.01).

Table 4.9 gives RB and MSE of coefficient estimates under various different hyper-hyper prior setting for r and δ. Accordingly, RB and MSE of coefficient posterior estimates are much less sensitive to the base setting of EB-GS hyper-prior for moderate sample sizes such as n=250.

Table 4.10 presents the performance measures for variable selection. Accordingly, overall, EB-GS type hyper-prior in Bayesian lasso for linear-circular regression is not sensitive to hyper-hyper prior choice for r and δ, for both sample sizes considered.

Table 4.9: Relative Bias (MSE) for Scenario I.

n	100		250	
Coefficients	Ga(0.01,0.01)	Ga(0.1,0.01)	Ga(0.01,0.01)	Ga(0.1,0.01)
β_0	0.42(6.54)	0.40(8.23)	0.28(4.27)	0.24(4.94)
β_1	-0.15(0.86)	-0.15(0.96)	-0.06(0.27)	-0.06(0.28)
β_2	-0.04(0.30)	-0.05(0.34)	-0.03(0.12)	-0.03(0.12)
β_3	-0.43(1.25)	-0.40(1.46)	-0.19(0.84)	-0.17(0.91)
β_4	-0.13(0.43)	-0.14(0.47)	-0.09(0.20)	-0.09(0.20)
β_5	NA(1.34)	NA(1.73)	NA(0.67)	NA(0.83)
β_6	NA(0.39)	NA(0.40)	NA(0.17)	NA(0.17)
β_7	-0.64(2.89)	-0.60(3.38)	-0.42(2.42)	-0.40(2.53)
β_8	-0.18(0.83)	-0.18(0.91)	-0.08(0.36)	-0.08(0.38)
β_9	NA(1.21)	NA(1.91)	NA(0.99)	NA(1.25)
β_{10}	NA(0.51)	NA(0.54)	NA(0.26)	NA(0.28)

NA: No results are available for these parameters.

Table 4.10: Performances in variable selection under Scenario I.

n	100		250	
Measures	Ga(0.01,0.01)	Ga(0.1,0.01)	Ga(0.01,0.01)	Ga(0.1,0.01)
ACI	1.959	1.962	2.841	2.833
AFE	1.041	1.038	0.159	0.167
CIR	0.254	0.264	0.853	0.841

4.1.2.4 Performance of EB-GS

In this section, we compare the performances of EB-GS with standard hyper-prior distributions given in the previous section. That is, we consider the following 3 hyper-prior distributions; *i.* Ga($\hat{r}$, $\hat{\delta}$), *ii.* Ga(0.1, 0.1) and *iii.* HC(0, 1). Note that Ga(0.1, 0.1) and HC(0, 1) outperformed the others as seen in Section 4.1.2.2 and thus EB-GS is compared against those. In this simulation study, hyper-hyper priors for r and δ are specified as $Ga(0.01, 0.01)$. All other details are the same as the Section 4.1.2.

In Ga($\hat{r}$, $\hat{\delta}$), $\hat{r}$ and $\hat{\delta}$ are posterior modes that are obtained by using proposed approach presented in Section 3.2.

4.1.2.5 Results

In order to investigate the performance of the Bayesian lasso with different hyper-prior distributions for the tuning parameter in terms of parameter estimations, RB and MSE are employed. Tables 4.11-4.16 show RBs and MSEs for this simulation study.

Overall, it can be observed that RBs and MSEs exhibit a pattern of comparability or reduction when the EB-GS hyper-prior distribution is employed in contrast to other hyper-prior distributions. This trend is particularly pronounced, and its effects are more prominent when $\kappa \geq 2$. Such findings suggest that the Bayesian lasso utilizing the EB-GS hyper-prior for tuning parameter selection demonstrates an enhanced level of performance in terms of the final parameter estimations within the context of linear-circular regression models. Additionally, it is noteworthy that RBs and MSEs display smaller values across all methods when $\kappa < 2$. This pattern underscores the consistent advantage of employing this specific parameter range across various methods, emphasizing its effectiveness in yielding improved results in terms of bias and mean squared error metrics.

Tables 4.17 and 4.18 show the variable selection measures for comparison of the Bayesian lasso with different hyper-prior distributions for the tuning parameter. In general, the variable selection measures when EB-GS hyper-prior distribution is em-

ployed consistently exhibit comparability or even superiority when compared to other hyper-prior distributions. This observation highlights the effectiveness of the proposed strategy in providing a more distinct differentiation between covariates that hold significance and those that do not. This distinct behavior implies that employing the Bayesian lasso with the EB-GS hyper-prior support modeling, within an appropriate dimensional framework that encompasses only the pertinent covariates, results in the identification of the optimal model with heightened precision as the sample size increases. Additionally, results show that the variable selection measures are superior when $\kappa < 2$ for all methods.

Table 4.11: Relative Bias (MSE) for Scenario I (When $\kappa \geq 2$).

n	100			250		
Coefficients	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)
β_0	0.36(6.05)	0.35(6.01)	0.39(6.14)	0.31(5.11)	0.30(5.07)	0.32(5.08)
β_1	-0.12(0.78)	-0.11(0.76)	-0.13(0.81)	-0.07(0.29)	-0.06(0.28)	-0.07(0.29)
β_2	-0.05(0.33)	-0.05(0.32)	-0.05(0.34)	-0.03(0.13)	-0.03(0.13)	-0.03(0.13)
β_3	-0.34(1.50)	-0.34(1.50)	-0.37(1.47)	-0.26(0.88)	-0.26(0.87)	-0.27(0.87)
β_4	-0.18(0.49)	-0.17(0.49)	-0.19(0.50)	-0.08(0.21)	-0.08(0.21)	-0.09(0.21)
β_5	NA(1.33)	NA(1.37)	NA(1.20)	NA(0.93)	NA(0.94)	NA(0.88)
β_6	NA(0.44)	NA(0.44)	NA(0.42)	NA(0.17)	NA(0.17)	NA(0.16)
β_7	-0.58(2.82)	-0.57(2.83)	-0.60(2.79)	-0.49(2.25)	-0.49(2.23)	-0.52(2.24)
β_8	-0.23(0.88)	-0.23(0.87)	-0.25(0.90)	-0.10(0.37)	-0.10(0.37)	-0.11(0.38)
β_9	NA(1.21)	NA(1.24)	NA(1.06)	NA(1.08)	NA(1.09)	NA(0.97)
β_{10}	NA(0.48)	NA(0.49)	NA(0.45)	NA(0.29)	NA(0.29)	NA(0.28)

NA: No results are available for these parameters.

Table 4.12: Relative Bias (MSE) for Scenario II (When $\kappa \geq 2$).

n	100			250		
Coefficients	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)
β_0	0.27(2.56)	0.24(2.46)	0.33(2.73)	0.23(2.15)	0.22(2.26)	0.28(2.32)
β_1	-0.29(0.40)	-0.26(0.42)	-0.42(0.40)	-0.20(0.19)	-0.20(0.20)	-0.30(0.22)
β_2	-0.24(0.23)	-0.22(0.23)	-0.35(0.27)	-0.11(0.10)	-0.11(0.10)	-0.18(0.12)
β_3	-0.64(0.72)	-0.59(0.80)	-0.71(0.69)	-0.41(0.55)	-0.39(0.56)	-0.51(0.54)
β_4	-0.30(0.32)	-0.27(0.33)	-0.42(0.35)	-0.17(0.16)	-0.16(0.16)	-0.26(0.19)
β_5	NA(0.36)	NA(0.47)	NA(0.25)	NA(0.34)	NA(0.37)	NA(0.23)
β_6	NA(0.19)	NA(0.21)	NA(0.13)	NA(0.14)	NA(0.14)	NA(0.11)
β_7	-0.75(0.89)	-0.71(0.80)	-0.82(0.72)	-0.66(0.71)	-0.63(0.79)	-0.74(0.68)
β_8	-0.43(0.48)	-0.40(0.50)	-0.55(0.48)	-0.26(0.25)	-0.25(0.26)	-0.36(0.28)
β_9	NA(0.34)	NA(0.34)	NA(0.17)	NA(0.25)	NA(0.36)	NA(0.18)
β_{10}	NA(0.33)	NA(0.38)	NA(0.23)	NA(0.19)	NA(0.20)	NA(0.14)

NA: No results are available for these parameters.

Table 4.13: Relative Bias (MSE) for Scenario III (When $\kappa \geq 2$).

n	100			250		
Coefficients	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)
β_0	1.74(59.04)	1.69(56.02)	1.7(56.77)	1.25(35.03)	1.23(34.04)	1.24(34.44)
β_1	-0.13(0.87)	-0.13(0.86)	-0.13(0.86)	-0.05(0.25)	-0.05(0.25)	-0.05(0.25)
β_2	-0.08(0.42)	-0.08(0.42)	-0.08(0.42)	-0.02(0.12)	-0.02(0.12)	-0.02(0.12)
β_3	-0.37(2.86)	-0.36(2.83)	-0.36(2.84)	-0.21(1.37)	-0.20(1.36)	-0.21(1.36)
β_4	-0.10(0.64)	-0.10(0.63)	-0.10(0.63)	-0.03(0.17)	-0.03(0.17)	-0.03(0.17)
β_5	-0.49(4.09)	-0.48(4.06)	-0.48(4.09)	-0.31(2.55)	-0.30(2.53)	-0.31(2.54)
β_6	-0.17(0.93)	-0.16(0.92)	-0.16(0.92)	-0.05(0.30)	-0.05(0.29)	-0.05(0.29)
β_7	-0.58(4.50)	-0.57(4.46)	-0.57(4.48)	-0.43(3.64)	-0.42(3.62)	-0.42(3.62)
β_8	-0.20(1.31)	-0.19(1.29)	-0.19(1.29)	-0.06(0.38)	-0.06(0.38)	-0.06(0.38)
β_9	-0.13(0.81)	-0.13(0.80)	-0.13(0.80)	-0.06(0.29)	-0.06(0.29)	-0.06(0.29)
β_{10}	-0.07(0.37)	-0.07(0.37)	-0.07(0.37)	-0.02(0.12)	-0.02(0.12)	-0.02(0.12)
β_{11}	-0.32(2.71)	-0.31(2.69)	-0.31(2.70)	-0.19(1.34)	-0.18(1.33)	-0.19(1.33)
β_{12}	-0.13(0.68)	-0.13(0.67)	-0.13(0.67)	-0.04(0.19)	-0.04(0.19)	-0.04(0.19)
β_{13}	-0.48(3.98)	-0.47(3.95)	-0.47(3.96)	-0.34(2.49)	-0.33(2.46)	-0.33(2.47)
β_{14}	-0.16(1.03)	-0.16(1.02)	-0.16(1.02)	-0.06(0.30)	-0.06(0.30)	-0.06(0.30)
β_{15}	-0.61(4.95)	-0.60(4.92)	-0.60(4.92)	-0.49(3.69)	-0.48(3.67)	-0.49(3.67)
β_{16}	-0.19(1.30)	-0.19(1.28)	-0.19(1.28)	-0.09(0.43)	-0.09(0.42)	-0.09(0.42)
β_{17}	NA(0.37)	NA(0.38)	NA(0.38)	NA(0.17)	NA(0.17)	NA(0.17)
β_{18}	NA(0.21)	NA(0.21)	NA(0.21)	NA(0.09)	NA(0.09)	NA(0.09)
β_{19}	NA(0.94)	NA(0.98)	NA(0.98)	NA(0.53)	NA(0.53)	NA(0.54)
β_{20}	NA(0.31)	NA(0.32)	NA(0.32)	NA(0.14)	NA(0.14)	NA(0.14)
β_{21}	NA(1.03)	NA(1.09)	NA(1.09)	NA(0.78)	NA(0.80)	NA(0.80)
β_{22}	NA(0.40)	NA(0.41)	NA(0.41)	NA(0.17)	NA(0.17)	NA(0.17)
β_{23}	NA(1.25)	NA(1.31)	NA(1.31)	NA(1.06)	NA(1.09)	NA(1.09)
β_{24}	NA(0.45)	NA(0.46)	NA(0.46)	NA(0.22)	NA(0.22)	NA(0.22)
β_{25}	NA(0.43)	NA(0.44)	NA(0.44)	NA(0.17)	NA(0.17)	NA(0.17)
β_{26}	NA(0.22)	NA(0.23)	NA(0.23)	NA(0.09)	NA(0.09)	NA(0.09)
β_{27}	NA(0.85)	NA(0.88)	NA(0.88)	NA(0.48)	NA(0.49)	NA(0.49)
β_{28}	NA(0.30)	NA(0.30)	NA(0.30)	NA(0.14)	NA(0.14)	NA(0.14)
β_{29}	NA(1.24)	NA(1.29)	NA(1.28)	NA(0.78)	NA(0.80)	NA(0.80)
β_{30}	NA(0.40)	NA(0.40)	NA(0.40)	NA(0.18)	NA(0.19)	NA(0.19)
β_{31}	NA(1.21)	NA(1.28)	NA(1.27)	NA(0.92)	NA(0.96)	NA(0.95)
β_{32}	NA(0.53)	NA(0.54)	NA(0.54)	NA(0.20)	NA(0.20)	NA(0.20)

NA: No results are available for these parameters.

Table 4.14: Relative Bias (MSE) for Scenario I (When $\kappa < 2$).

n	100			250		
Coefficients	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)
β_0	0.09(0.76)	0.09(0.75)	0.10(0.77)	0.03(0.26)	0.03(0.26)	0.03(0.26)
β_1	-0.06(0.25)	-0.06(0.25)	-0.06(0.26)	-0.02(0.08)	-0.02(0.08)	-0.02(0.08)
β_2	-0.05(0.23)	-0.05(0.22)	-0.06(0.23)	-0.01(0.08)	-0.01(0.08)	-0.02(0.08)
β_3	-0.11(0.27)	-0.11(0.27)	-0.12(0.28)	-0.03(0.11)	-0.03(0.11)	-0.04(0.11)
β_4	-0.1(0.25)	-0.10(0.24)	-0.11(0.25)	-0.03(0.08)	-0.03(0.08)	-0.04(0.08)
β_5	NA(0.23)	NA(0.23)	NA(0.22)	NA(0.10)	NA(0.10)	NA(0.10)
β_6	NA(0.16)	NA(0.16)	NA(0.15)	NA(0.07)	NA(0.07)	NA(0.07)
β_7	-0.17(0.58)	-0.17(0.57)	-0.18(0.59)	-0.06(0.19)	-0.06(0.19)	-0.06(0.19)
β_8	-0.10(0.30)	-0.10(0.30)	-0.11(0.31)	-0.04(0.10)	-0.04(0.10)	-0.04(0.10)
β_9	NA(0.32)	NA(0.32)	NA(0.30)	NA(0.17)	NA(0.17)	NA(0.17)
β_{10}	NA(0.17)	NA(0.17)	NA(0.17)	NA(0.08)	NA(0.08)	NA(0.08)

NA: No results are available for these parameters.

Table 4.15: Relative Bias (MSE) for Scenario II (When $\kappa < 2$).

n	100			250		
Coefficients	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)
β_0	0.07(0.49)	0.07(0.50)	0.11(0.55)	0.03(0.24)	0.04(0.24)	0.06(0.26)
β_1	-0.22(0.20)	-0.22(0.20)	-0.33(0.23)	-0.11(0.09)	-0.12(0.09)	-0.17(0.10)
β_2	-0.18(0.19)	-0.18(0.19)	-0.30(0.22)	-0.12(0.08)	-0.12(0.08)	-0.17(0.09)
β_3	-0.22(0.23)	-0.22(0.23)	-0.34(0.26)	-0.12(0.10)	-0.13(0.11)	-0.19(0.12)
β_4	-0.21(0.21)	-0.21(0.21)	-0.32(0.24)	-0.12(0.09)	-0.13(0.09)	-0.18(0.10)
β_5	NA(0.19)	NA(0.19)	NA(0.13)	NA(0.09)	NA(0.08)	NA(0.07)
β_6	NA(0.14)	NA(0.14)	NA(0.10)	NA(0.06)	NA(0.06)	NA(0.05)
β_7	-0.31(0.33)	-0.31(0.34)	-0.45(0.35)	-0.16(0.17)	-0.17(0.18)	-0.24(0.19)
β_8	-0.23(0.23)	-0.23(0.23)	-0.34(0.26)	-0.14(0.10)	-0.15(0.10)	-0.20(0.12)
β_9	NA(0.21)	NA(0.22)	NA(0.14)	NA(0.12)	NA(0.11)	NA(0.09)
β_{10}	NA(0.15)	NA(0.16)	NA(0.11)	NA(0.07)	NA(0.07)	NA(0.06)

NA: No results are available for these parameters.

Table 4.16: Relative Bias (MSE) for Scenario III (When $\kappa < 2$).

n	100			250		
Coefficients	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)
β_0	0.34(4.43)	0.33(4.34)	0.33(4.36)	0.10(0.8)	0.10(0.79)	0.10(0.80)
β_1	-0.07(0.34)	-0.06(0.34)	-0.06(0.34)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)
β_2	-0.04(0.27)	-0.04(0.27)	-0.04(0.27)	-0.01(0.09)	-0.01(0.08)	-0.01(0.09)
β_3	-0.08(0.41)	-0.08(0.41)	-0.08(0.41)	-0.02(0.12)	-0.02(0.12)	-0.02(0.12)
β_4	-0.05(0.31)	-0.05(0.31)	-0.05(0.31)	-0.02(0.09)	-0.02(0.09)	-0.02(0.09)
β_5	-0.11(0.67)	-0.10(0.66)	-0.10(0.66)	-0.04(0.20)	-0.03(0.20)	-0.04(0.20)
β_6	-0.07(0.38)	-0.07(0.37)	-0.07(0.37)	-0.03(0.12)	-0.03(0.12)	-0.03(0.12)
β_7	-0.16(1.02)	-0.15(1.01)	-0.15(1.01)	-0.04(0.23)	-0.04(0.23)	-0.04(0.23)
β_8	-0.08(0.46)	-0.08(0.46)	-0.08(0.46)	-0.03(0.11)	-0.03(0.11)	-0.03(0.11)
β_9	-0.06(0.32)	-0.06(0.32)	-0.06(0.32)	-0.02(0.09)	-0.02(0.09)	-0.02(0.09)
β_{10}	-0.04(0.29)	-0.04(0.29)	-0.04(0.29)	-0.02(0.09)	-0.02(0.09)	-0.01(0.09)
β_{11}	-0.07(0.42)	-0.07(0.42)	-0.07(0.42)	-0.03(0.12)	-0.03(0.12)	-0.03(0.12)
β_{12}	-0.05(0.31)	-0.05(0.31)	-0.05(0.31)	-0.01(0.08)	-0.01(0.08)	-0.01(0.08)
β_{13}	-0.09(0.64)	-0.09(0.63)	-0.09(0.63)	-0.04(0.17)	-0.04(0.17)	-0.04(0.17)
β_{14}	-0.06(0.35)	-0.05(0.35)	-0.05(0.35)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)
β_{15}	-0.17(1.05)	-0.16(1.04)	-0.16(1.04)	-0.05(0.23)	-0.05(0.23)	-0.05(0.23)
β_{16}	-0.09(0.44)	-0.09(0.44)	-0.09(0.44)	-0.02(0.10)	-0.02(0.10)	-0.02(0.10)
β_{17}	NA(0.20)	NA(0.20)	NA(0.20)	NA(0.08)	NA(0.08)	NA(0.08)
β_{18}	NA(0.19)	NA(0.19)	NA(0.19)	NA(0.07)	NA(0.07)	NA(0.07)
β_{19}	NA(0.25)	NA(0.26)	NA(0.26)	NA(0.08)	NA(0.08)	NA(0.08)
β_{20}	NA(0.23)	NA(0.23)	NA(0.23)	NA(0.07)	NA(0.07)	NA(0.07)
β_{21}	NA(0.33)	NA(0.33)	NA(0.33)	NA(0.12)	NA(0.12)	NA(0.12)
β_{22}	NA(0.25)	NA(0.25)	NA(0.25)	NA(0.09)	NA(0.09)	NA(0.09)
β_{23}	NA(0.44)	NA(0.45)	NA(0.45)	NA(0.13)	NA(0.13)	NA(0.13)
β_{24}	NA(0.25)	NA(0.26)	NA(0.26)	NA(0.08)	NA(0.08)	NA(0.08)
β_{25}	NA(0.18)	NA(0.18)	NA(0.18)	NA(0.08)	NA(0.08)	NA(0.08)
β_{26}	NA(0.20)	NA(0.20)	NA(0.20)	NA(0.06)	NA(0.06)	NA(0.06)
β_{27}	NA(0.23)	NA(0.23)	NA(0.23)	NA(0.11)	NA(0.11)	NA(0.11)
β_{28}	NA(0.19)	NA(0.19)	NA(0.19)	NA(0.08)	NA(0.08)	NA(0.08)
β_{29}	NA(0.37)	NA(0.38)	NA(0.38)	NA(0.11)	NA(0.11)	NA(0.11)
β_{30}	NA(0.22)	NA(0.22)	NA(0.22)	NA(0.09)	NA(0.09)	NA(0.09)
β_{31}	NA(0.48)	NA(0.49)	NA(0.49)	NA(0.18)	NA(0.18)	NA(0.18)
β_{32}	NA(0.28)	NA(0.29)	NA(0.29)	NA(0.08)	NA(0.08)	NA(0.08)

NA: No results are available for these parameters.

Table 4.17: Performances in variable selection under Scenario I-III (When $\kappa \geq 2$).

		Scenario I (ACI=3, AFE=0, CIR=1)*			Scenario II(ACI=3, AFE=0, CIR=1)*			Scenario III(ACI=8, AFE=0, CIR=1)*		
n	Measures	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)
	ACI	1.850	1.910	1.884	0.514	0.520	0.296	6.960	6.914	6.952
100	AFE	1.150	1.090	1.116	2.486	2.480	2.704	1.040	1.086	1.048
	CIR	0.200	0.224	0.216	0.002	0.004	0.000	0.322	0.314	0.324
	ACI	2.805	2.812	2.812	1.403	1.417	1.098	8.000	8.000	8.000
250	AFE	0.195	0.188	0.188	1.597	1.583	1.902	0.000	0.000	0.000
	CIR	0.815	0.819	0.819	0.068	0.077	0.041	1.000	1.000	1.000

*: The true values of variable selection measures.

Table 4.18: Performances in variable selection under Scenario I-III (When $\kappa < 2$).

		Scenario I (ACI=3, AFE=0, CIR=1)*			Scenario II (ACI=3, AFE=0, CIR=1)*			Scenario III (ACI=8, AFE=0, CIR=1)*		
n	Measures	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)	Ga(0.1, 0.1)	Ga($\hat{r}$, $\hat{\delta}$)	HC(0, 1)
	ACI	2.926	2.918	2.924	1.396	1.426	0.918	7.998	7.998	7.998
100	AFE	0.074	0.082	0.076	1.604	1.574	2.082	0.002	0.002	0.002
	CIR	0.926	0.918	0.924	0.102	0.102	0.058	0.998	0.998	0.998
	ACI	3.000	3.000	3.000	2.658	2.686	2.514	8.000	8.000	8.000
250	AFE	0.000	0.000	0.000	0.342	0.314	0.486	0.000	0.000	0.000
	CIR	1.000	1.000	1.000	0.686	0.712	0.592	1.000	1.000	1.000

*: The true values of variable selection measures.

4.1.3 Simulation Study: Bayesian Lasso in Circular-(Circular, Linear) Regression Models

This section describes a simulation design evaluating the accuracy of parameter estimation and variable selection performance of the proposed methodologies in Section 3.3.2.

4.1.3.1 Simulation Design

Simulation design is controlled for the following factors:

- Sample size (n): The simulation is performed for three different sample sizes

- 50, 100, and 250. This allows us to examine finite sample properties of the methods, i.e. how analysis and the stability of the results are affected by sample size.

- Size of regression coefficients (β): The regression coefficients are varied within a specific range, from 0.85 to 4. This variation helps us understand the effect of different coefficient magnitudes in the underlying data generating mechanism on the method's performance and estimation.

- Concentration parameter of the circular dependent variable (κ): The concentration parameter is categorized into two groups: $\kappa < 2(\kappa = 1)$ and $\kappa \geq 2(\kappa = 3)$. By considering these values, we can investigate how the concentration of the data affects the analysis.

- Number of independent variables: The number of independent variables in simulation varies between 8 and 16. This aspect allows us to explore the impact of the number of predictors on the analysis.

By controlling these factors in our simulation, we systematically investigate the results of statistical analysis. This approach helps to ensure that conclusions are robust and generalizable across different scenarios.

In all simulation settings, circular covariates are independently generated from the von Mises distributions with circular mean 0 without loss of generality and various different concentration parameters, $\kappa = 2, 4, 6, 8, 10$, while linear covariates are independently generated from the standard normal distribution.

The four simulation scenarios considered are as follows:

- **Scenario I**: $\beta = (\underbrace{3, 4, 0, 0}_{\text{Circular}}, \underbrace{3, 4, 0, 0}_{\text{Linear}})$ p=4, q=4
- **Scenario II**: $\beta = (\underbrace{0.85, 0.85, 0, 0}_{\text{Circular}}, \underbrace{0.85, 0.85, 0, 0}_{\text{Linear}})$, p=4, q=4
- **Scenario III**: $\beta = (\underbrace{3, 4, 3, 4, 0, 0, 0, 0}_{\text{Circular}}, \underbrace{3, 4, 3, 4, 0, 0, 0, 0}_{\text{Linear}})$, p=8, q=8
- **Scenario IV**: $\beta = (\underbrace{0.85, 0.85, 0.85, 0.85, 0, 0, 0, 0}_{\text{Circular}}, \underbrace{0.85, 0.85, 0.85, 0.85, 0, 0, 0, 0}_{\text{Linear}})$, p=8, q=8.

where p and q are the number of circular and linear covariates, respectively. μ_θ is set to 1.5 in all scenarios. A gamma prior (Gamma(0.01,0.01)) is used as prior distribution for the tuning parameter in Classical Bayesian lasso and BaLasso. G_0 is specified as Gamma(0.01,0.01) in $DLasso_1$ and $DLasso_2$. In Dirichlet mixture lasso, G_{01} and G_{02} are specified as Gamma(0.01, 0.01) and Gamma(0.1, 0.1), respectively. Each scenario is repeated 100 times.

4.1.3.2 Results

The simulation study evaluates the performance of the approaches in terms of parameter estimations using two metrics: RB and MSE. The results are presented in Tables 4.19 to 4.42. In these tables, the methods given in Section 3.3.2 are represented by $M_1, M_2, ..., M_5$, respectively.

The results indicate that the different methods being compared in this study have similar performance based on two evaluation metrics for parameter estimation. In other words, there is no clear evidence that one method significantly outperforms the others in terms of bias and accuracy of parameter estimations for all scenarios. On the other hand, the use of the Bayesian lasso methods with flexible priors for λ seems to have improved the performance of parameter estimations for some scenarios where $\kappa \geq 2$ and n is relatively small (50 and 100).

Variable selection measures are used to assess the performances of different methods in selecting significant variables from a set of potential predictors in simulation

studies. The results are available in Tables 4.43 to 4.46. Upon a comprehensive assessment of the results, it's clear that the measures for variable selection are quite similar. Moreover, it is important to note that all approaches perform perfectly for a relatively larger sample size, with n=250.

Table 4.19: Relative Bias (MSE) for Scenario I, n=50, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.04(1.16)	0.02(1.46)	-0.01(1.21)	-0.02(1.08)	-0.01(1.25)
β_2	-0.05(1.30)	0.08(1.71)	-0.01(1.38)	0.02(1.10)	-0.01(1.40)
β_3	NA(0.61)	NA(0.53)	NA(0.63)	NA(0.34)	NA(0.65)
β_4	NA(0.42)	NA(0.20)	NA(0.43)	NA(0.09)	NA(0.43)
β_5	0.03(0.39)	0.11(0.54)	0.06(0.44)	0.06(0.38)	0.06(0.45)
β_6	0.04(0.64)	0.11(0.94)	0.06(0.73)	0.05(0.65)	0.07(0.74)
β_7	NA(0.05)	NA(0.03)	NA(0.05)	NA(0.01)	NA(0.05)
β_8	NA(0.06)	NA(0.03)	NA(0.06)	NA(0.00)	NA(0.06)
μ_θ	0.02(0.02)	0.01(0.02)	0.01(0.02)	0.01(0.02)	0.01(0.02)
κ	-0.03(0.20)	-0.02(0.20)	-0.03(0.20)	-0.02(0.19)	-0.03(0.20)

NA: No results are available for these parameters.

Table 4.20: Relative Bias (MSE) for Scenario I, n=100, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.04(0.43)	0.01(0.45)	-0.02(0.43)	-0.01(0.38)	-0.02(0.43)
β_2	-0.01(0.43)	0.04(0.48)	0.00(0.44)	0.01(0.39)	0.01(0.46)
β_3	NA(0.18)	NA(0.11)	NA(0.18)	NA(0.07)	NA(0.18)
β_4	NA(0.17)	NA(0.08)	NA(0.17)	NA(0.02)	NA(0.19)
β_5	0.02(0.12)	0.04(0.14)	0.02(0.12)	0.01(0.11)	0.02(0.13)
β_6	0.01(0.22)	0.04(0.26)	0.02(0.23)	0.02(0.22)	0.02(0.24)
β_7	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_8	NA(0.01)	NA(0.01)	NA(0.02)	NA(0.00)	NA(0.02)
μ_θ	0.01(0.01)	0.00(0.01)	0.01(0.01)	0.01(0.01)	0.01(0.01)
κ	-0.01(0.18)	0.00(0.19)	-0.01(0.18)	0.00(0.19)	-0.01(0.18)

NA: No results are available for these parameters.

Table 4.21: Relative Bias (MSE) for Scenario I, n=250, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.01(0.12)	0.01(0.12)	0.00(0.12)	-0.01(0.10)	0.00(0.13)
β_2	0.00(0.15)	0.02(0.16)	0.01(0.16)	0.00(0.15)	0.01(0.16)
β_3	NA(0.07)	NA(0.04)	NA(0.06)	NA(0.01)	NA(0.07)
β_4	NA(0.06)	NA(0.03)	NA(0.05)	NA(0.00)	NA(0.06)
β_5	0.00(0.05)	0.01(0.05)	0.01(0.05)	0.00(0.05)	0.01(0.05)
β_6	0.00(0.09)	0.02(0.09)	0.01(0.09)	0.00(0.08)	0.01(0.09)
β_7	NA(0.01)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.01)
β_8	NA(0.01)	NA(0.00)	NA(0.01)	NA(0.00)	NA(0.01)
μ_θ	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.00(0.00)
κ	0.00(0.05)	0.00(0.05)	-0.01(0.05)	0.00(0.05)	0.00(0.05)

NA: No results are available for these parameters.

Table 4.22: Relative Bias (MSE) for Scenario II, n=50, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	0.15(0.13)	-0.03(0.17)	-0.09(0.13)	-0.13(0.19)	-0.07(0.13)
β_2	-0.15(0.13)	-0.01(0.18)	-0.08(0.13)	-0.11(0.19)	-0.06(0.13)
β_3	NA(0.04)	NA(0.04)	NA(0.05)	NA(0.02)	NA(0.05)
β_4	NA(0.06)	NA(0.06)	NA(0.07)	NA(0.03)	NA(0.07)
β_5	0.01(0.02)	0.07(0.03)	0.04(0.02)	0.03(0.02)	0.05(0.02)
β_6	0.02(0.02)	0.07(0.03)	0.05(0.03)	0.04(0.02)	0.06(0.03)
β_7	NA(0.01)	NA(0.00)	NA(0.01)	NA(0.00)	NA(0.01)
β_8	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
μ_θ	0.04(0.04)	0.01(0.03)	0.02(0.03)	0.04(0.04)	0.02(0.03)
κ	-0.04(0.23)	-0.03(0.23)	-0.04(0.23)	-0.04(0.24)	-0.04(0.23)

NA: No results are available for these parameters.

Table 4.23: Relative Bias (MSE) for Scenario II, n=100, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.11(0.07)	-0.05(0.08)	-0.08(0.07)	-0.07(0.09)	-0.07(0.07)
β_2	-0.06(0.05)	0.01(0.06)	-0.03(0.05)	0.00(0.06)	-0.02(0.05)
β_3	NA(0.03)	NA(0.02)	NA(0.03)	NA(0.01)	NA(0.03)
β_4	NA(0.03)	NA(0.02)	NA(0.03)	NA(0.01)	NA(0.03)
β_5	0.02(0.01)	0.04(0.01)	0.03(0.01)	0.03(0.01)	0.03(0.01)
β_6	0.02(0.01)	0.04(0.01)	0.03(0.01)	0.03(0.01)	0.03(0.01)
β_7	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
β_8	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
μ_θ	0.03(0.02)	0.01(0.02)	0.02(0.02)	0.02(0.02)	0.02(0.02)
κ	-0.04(0.12)	-0.04(0.12)	-0.04(0.12)	-0.04(0.11)	-0.04(0.12)

NA: No results are available for these parameters.

Table 4.24: Relative Bias (MSE) for Scenario II, n=250, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	0.04(0.04)	0.07(0.04)	0.05(0.04)	0.06(0.04)	0.05(0.04)
β_2	0.00(0.02)	0.03(0.02)	0.01(0.02)	0.02(0.02)	0.01(0.02)
β_3	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_4	NA(0.02)	NA(0.02)	NA(0.02)	NA(0.00)	NA(0.02)
β_5	0.01(0.00)	0.02(0.00)	0.02(0.00)	0.02(0.00)	0.02(0.00)
β_6	0.02(0.00)	0.03(0.00)	0.02(0.00)	0.02(0.00)	0.02(0.00)
β_7	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
β_8	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
μ_θ	0.00(0.01)	-0.01(0.01)	0.00(0.01)	-0.01(0.01)	0.00(0.01)
κ	-0.01(0.06)	-0.01(0.06)	-0.01(0.06)	-0.01(0.06)	-0.01(0.06)

NA: No results are available for these parameters.

Table 4.25: Relative Bias (MSE) for Scenario III, n=50, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.68(8.31)	-0.28(5.71)	-0.8(10.92)	-0.37(6.31)	-0.54(4.28)
β_2	-0.80(16.14)	-0.20(4.85)	-0.72(16.97)	-0.27(3.58)	-0.52(6.85)
β_3	-0.74(9.77)	-0.50(3.93)	-0.73(11.45)	-0.57(4.05)	-0.59(4.58)
β_4	-0.74(13.27)	-0.29(6.18)	-0.69(16.69)	-0.36(5.72)	-0.63(8.27)
β_5	NA(2.63)	NA(0.50)	NA(5.18)	NA(0.15)	NA(1.23)
β_6	NA(2.98)	NA(1.42)	NA(4.71)	NA(0.22)	NA(1.88)
β_7	NA(4.83)	NA(0.99)	NA(3.88)	NA(0.33)	NA(1.41)
β_8	NA(3.11)	NA(1.06)	NA(3.99)	NA(0.80)	NA(1.94)
β_9	-0.60(9.91)	-0.14(0.36)	-0.64(13.76)	-0.22(1.05)	-0.41(3.45)
β_{10}	-0.52(10.64)	-0.18(0.97)	-0.5(17.20)	-0.28(2.44)	-0.41(5.08)
β_{11}	-0.71(11.95)	-0.20(0.93)	-0.56(11.73)	-0.27(1.62)	-0.42(3.09)
β_{12}	-0.50(11.66)	-0.17(1.07)	-0.74(22.55)	-0.22(1.94)	-0.37(4.84)
β_{13}	NA(7.03)	NA(0.12)	NA(6.89)	NA(0.16)	NA(1.24)
β_{14}	NA(5.26)	NA(0.07)	NA(6.39)	NA(5.28)	NA(1.66)
β_{15}	NA(6.85)	NA(0.04)	NA(4.86)	NA(0.31)	NA(1.11)
β_{16}	NA(6.25)	NA(0.10)	NA(9.22)	NA(0.20)	NA(0.61)
μ_θ	-0.06(0.04)	0.07(0.03)	-0.07(0.05)	0.07(0.03)	0.00(0.03)
κ	-0.16(0.68)	0.17(0.47)	-0.23(1.00)	0.13(0.35)	-0.06(0.46)

NA: No results are available for these parameters.

Table 4.26: Relative Bias (MSE) for Scenario III, n=100, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.12(0.97)	-0.03(1.43)	-0.11(0.96)	-0.11(0.91)	-0.11(0.98)
β_2	-0.17(1.73)	-0.03(1.84)	-0.14(1.64)	-0.10(1.40)	-0.14(1.79)
β_3	-0.22(1.28)	-0.15(1.53)	-0.21(1.24)	-0.18(1.38)	-0.21(1.24)
β_4	-0.21(2.14)	-0.08(2.24)	-0.19(2.04)	-0.12(1.58)	-0.19(2.11)
β_5	NA(0.54)	NA(0.28)	NA(0.57)	NA(0.14)	NA(0.57)
β_6	NA(0.52)	NA(0.37)	NA(0.53)	NA(0.20)	NA(0.54)
β_7	NA(0.73)	NA(0.47)	NA(0.74)	NA(0.20)	NA(0.76)
β_8	NA(0.56)	NA(0.50)	NA(0.55)	NA(0.34)	NA(0.58)
β_9	-0.11(0.30)	-0.02(0.25)	-0.10(0.29)	-0.09(0.23)	-0.09(0.28)
β_{10}	-0.10(0.51)	-0.01(0.51)	-0.09(0.46)	-0.08(0.41)	-0.08(0.48)
β_{11}	-0.11(0.30)	-0.03(0.23)	-0.09(0.28)	-0.09(0.24)	-0.09(0.29)
β_{12}	-0.10(0.47)	-0.02(0.34)	-0.09(0.43)	-0.09(0.36)	-0.08(0.45)
β_{13}	NA(0.05)	NA(0.02)	NA(0.05)	NA(0.00)	NA(0.05)
β_{14}	NA(0.06)	NA(0.03)	NA(0.05)	NA(0.00)	NA(0.06)
β_{15}	NA(0.06)	NA(0.04)	NA(0.06)	NA(0.01)	NA(0.06)
β_{16}	NA(0.05)	NA(0.02)	NA(0.04)	NA(0.00)	NA(0.05)
μ_θ	0.02(0.01)	0.01(0.01)	0.01(0.01)	0.01(0.01)	0.01(0.01)
κ	0.02(0.10)	0.02(0.10)	0.02(0.10)	0.02(0.10)	0.02(0.10)

NA: No results are available for these parameters.

Table 4.27: Relative Bias (MSE) for Scenario III, n=250, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.01(0.27)	0.04(0.31)	0.01(0.28)	-0.03(0.29)	-0.01(0.30)
β_2	0.01(0.28)	0.06(0.34)	0.02(0.29)	-0.06(0.68)	0.00(0.45)
β_3	0.01(0.37)	0.06(0.41)	0.02(0.38)	0.05(0.39)	-0.03(0.40)
β_4	0.00(0.25)	0.04(0.31)	0.01(0.27)	-0.02(0.25)	-0.01(0.30)
β_5	NA(0.12)	NA(0.06)	NA(0.10)	NA(0.14)	NA(0.11)
β_6	NA(0.14)	NA(0.07)	NA(0.13)	NA(0.04)	NA(0.15)
β_7	NA(0.15)	NA(0.07)	NA(0.13)	NA(0.46)	NA(0.14)
β_8	NA(0.11)	NA(0.05)	NA(0.09)	NA(0.35)	NA(0.16)
β_9	0.01(0.07)	0.05(0.10)	0.02(0.08)	-0.03(0.10)	0.00(0.08)
β_{10}	0.01(0.15)	0.05(0.21)	0.03(0.17)	-0.03(0.19)	0.00(0.13)
β_{11}	0.01(0.08)	0.05(0.11)	0.02(0.09)	-0.01(0.28)	0.00(0.08)
β_{12}	0.01(0.14)	0.05(0.19)	0.02(0.15)	-0.03(0.18)	0.00(0.14)
β_{13}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_{14}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.01)
β_{15}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.01)
β_{16}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
μ_θ	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.01(0.00)	0.01(0.00)
κ	0.01(0.05)	0.01(0.05)	0.01(0.05)	-0.01(0.06)	-0.02(0.05)

NA: No results are available for these parameters.

Table 4.28: Relative Bias (MSE) for Scenario IV, n=50, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.24(0.27)	0.02(0.54)	-0.18(0.29)	-0.20(0.35)	-0.15(0.31)
β_2	-0.28(0.19)	-0.03(0.34)	-0.21(0.19)	-0.25(0.29)	-0.19(0.20)
β_3	-0.23(0.21)	0.02(0.36)	-0.17(0.21)	-0.19(0.26)	-0.14(0.22)
β_4	-0.22(0.22)	0.00(0.39)	-0.16(0.22)	-0.18(0.30)	-0.14(0.23)
β_5	NA(0.08)	NA(0.10)	NA(0.09)	NA(0.05)	NA(0.10)
β_6	NA(0.09)	NA(0.13)	NA(0.10)	NA(0.06)	NA(0.11)
β_7	NA(0.10)	NA(0.17)	NA(0.12)	NA(0.08)	NA(0.13)
β_8	NA(0.07)	NA(0.11)	NA(0.09)	NA(0.05)	NA(0.10)
β_9	-0.03(0.04)	0.11(0.06)	0.01(0.04)	0.00(0.03)	0.02(0.04)
β_{10}	-0.06(0.02)	0.08(0.03)	-0.02(0.02)	-0.02(0.02)	-0.01(0.02)
β_{11}	-0.06(0.03)	0.08(0.05)	-0.02(0.03)	-0.02(0.03)	-0.01(0.03)
β_{12}	-0.04(0.03)	0.10(0.04)	0.00(0.03)	-0.01(0.03)	0.01(0.03)
β_{13}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_{14}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.02)
β_{15}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_{16}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.02)
μ_θ	0.07(0.05)	0.00(0.04)	0.05(0.04)	0.07(0.05)	0.04(0.04)
κ	-0.02(0.24)	-0.01(0.23)	-0.02(0.23)	-0.02(0.23)	-0.02(0.23)

NA: No results are available for these parameters.

Table 4.29: Relative Bias (MSE) for Scenario IV, n=100, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.15(0.13)	-0.04(0.16)	-0.12(0.13)	-0.11(0.16)	-0.11(0.13)
β_2	-0.09(0.08)	0.03(0.12)	-0.06(0.08)	-0.02(0.12)	-0.05(0.08)
β_3	-0.10(0.12)	0.01(0.17)	-0.07(0.12)	-0.07(0.16)	-0.06(0.12)
β_4	-0.14(0.12)	-0.04(0.15)	-0.11(0.12)	-0.12(0.15)	-0.10(0.12)
β_5	NA(0.06)	NA(0.06)	NA(0.06)	NA(0.03)	NA(0.07)
β_6	NA(0.05)	NA(0.05)	NA(0.06)	NA(0.02)	NA(0.06)
β_7	NA(0.04)	NA(0.04)	NA(0.04)	NA(0.01)	NA(0.05)
β_8	NA(0.07)	NA(0.07)	NA(0.07)	NA(0.03)	NA(0.08)
β_9	-0.03(0.01)	0.03(0.02)	-0.01(0.01)	-0.02(0.01)	0.00(0.01)
β_{10}	-0.02(0.01)	0.05(0.02)	0.00(0.01)	0.00(0.01)	0.01(0.01)
β_{11}	-0.03(0.01)	0.03(0.01)	-0.01(0.01)	-0.01(0.01)	-0.01(0.01)
β_{12}	-0.01(0.01)	0.05(0.02)	0.01(0.01)	0.00(0.01)	0.01(0.01)
β_{13}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_{14}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_{15}	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
β_{16}	NA(0.01)	NA(0.00)	NA(0.01)	NA(0.00)	NA(0.01)
μ_θ	0.03(0.02)	0(0.02)	0.02(0.02)	0.03(0.02)	0.02(0.02)
κ	-0.01(0.14)	-0.01(0.13)	-0.01(0.13)	-0.02(0.13)	-0.01(0.13)

NA: No results are available for these parameters.

Table 4.30: Relative Bias (MSE) for Scenario IV, n=250, $\kappa = 3$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.02(0.04)	0.03(0.04)	-0.01(0.04)	0.01(0.04)	0.00(0.04)
β_2	-0.03(0.03)	0.01(0.03)	-0.02(0.03)	0.00(0.03)	-0.02(0.03)
β_3	-0.02(0.04)	0.03(0.04)	-0.01(0.04)	0.02(0.03)	0.00(0.04)
β_4	-0.05(0.03)	0.00(0.04)	-0.04(0.03)	-0.02(0.03)	-0.03(0.03)
β_5	NA(0.02)	NA(0.01)	NA(0.02)	NA(0.00)	NA(0.02)
β_6	NA(0.02)	NA(0.02)	NA(0.02)	NA(0.01)	NA(0.02)
β_7	NA(0.02)	NA(0.02)	NA(0.02)	NA(0.01)	NA(0.02)
β_8	NA(0.02)	NA(0.02)	NA(0.02)	NA(0.00)	NA(0.02)
β_9	0.01(0.00)	0.03(0.00)	0.02(0.00)	0.02(0.00)	0.02(0.00)
β_{10}	0.01(0.00)	0.03(0.00)	0.01(0.00)	0.02(0.00)	0.02(0.00)
β_{11}	-0.01(0.00)	0.02(0.01)	0.00(0.00)	0.00(0.00)	0.00(0.00)
β_{12}	0.01(0.00)	0.03(0.01)	0.01(0.00)	0.02(0.00)	0.02(0.00)
β_{13}	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
β_{14}	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
β_{15}	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
β_{16}	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)	NA(0.00)
μ_θ	0.00(0.01)	-0.01(0.01)	0.00(0.01)	0.00(0.01)	0.00(0.01)
κ	-0.01(0.04)	-0.01(0.04)	-0.01(0.04)	-0.01(0.04)	-0.01(0.04)

NA: No results are available for these parameters.

Table 4.31: Relative Bias (MSE) for Scenario I, n=50, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.75(6.85)	-0.70(7.74)	-0.70(7.19)	-0.79(7.26)	-0.72(7.08)
β_2	-0.75(11.24)	-0.49(9.07)	-0.68(11.55)	-0.78(12.34)	-0.76(11.89)
β_3	NA(1.49)	NA(3.04)	NA(2.30)	NA(2.09)	NA(2.29)
β_4	NA(1.85)	NA(2.73)	NA(2.05)	NA(1.74)	NA(1.62)
β_5	-0.53(4.87)	-0.23(6.92)	-0.41(5.27)	-0.71(7.19)	-0.5(4.86)
β_6	-0.44(6.87)	-0.24(6.85)	-0.39(7.62)	-0.57(8.34)	-0.47(7.85)
β_7	NA(1.61)	NA(0.85)	NA(2.52)	NA(2.36)	NA(2.01)
β_8	NA(1.65)	NA(0.32)	NA(3.23)	NA(2.02)	NA(1.16)
μ_θ	-0.09(0.24)	0.02(0.09)	-0.04(0.12)	-0.15(0.4)	-0.01(0.11)
κ	0.14(0.11)	0.23(0.14)	0.15(0.11)	0.06(0.07)	0.15(0.10)

NA: No results are available for these parameters.

Table 4.32: Relative Bias (MSE) for Scenario I, n=100, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.34(3.35)	-0.03(3.98)	-0.28(3.73)	-0.36(3.81)	-0.42(3.76)
β_2	-0.41(4.53)	-0.18(4.72)	-0.37(4.55)	-0.34(4.90)	-0.41(4.68)
β_3	NA(1.98)	NA(3.37)	NA(2.35)	NA(2.05)	NA(4.90)
β_4	NA(1.74)	NA(2.78)	NA(2.09)	NA(1.34)	NA(2.93)
β_5	-0.15(1.12)	-0.01(1.21)	-0.11(1.08)	-0.14(1.24)	-0.16(1.60)
β_6	-0.11(1.80)	0.08(1.99)	-0.06(1.64)	-0.09(1.74)	-0.12(1.86)
β_7	NA(0.27)	NA(0.28)	NA(0.35)	NA(0.18)	NA(0.99)
β_8	NA(0.47)	NA(0.29)	NA(0.71)	NA(0.37)	NA(0.76)
μ_θ	0.02(0.04)	0.02(0.03)	0.02(0.03)	0.00(0.11)	0.00(0.04)
κ	0.09(0.05)	0.10(0.05)	0.09(0.05)	0.08(0.05)	0.06(0.05)

NA: No results are available for these parameters.

Table 4.33: Relative Bias (MSE) for Scenario I, n=250, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.07(0.87)	0.00(1.08)	-0.05(0.89)	-0.03(0.97)	-0.05(0.92)
β_2	-0.04(1.43)	0.06(1.77)	0.00(1.53)	0.01(1.29)	0.00(1.54)
β_3	NA(0.38)	NA(0.25)	NA(0.40)	NA(0.12)	NA(0.42)
β_4	NA(0.37)	NA(0.19)	NA(0.38)	NA(0.09)	NA(0.39)
β_5	-0.01(0.33)	0.06(0.41)	0.02(0.35)	0.02(0.34)	0.02(0.36)
β_6	-0.01(0.57)	0.06(0.74)	0.01(0.61)	0.02(0.61)	0.02(0.63)
β_7	NA(0.04)	NA(0.02)	NA(0.05)	NA(0.01)	NA(0.05)
β_8	NA(0.04)	NA(0.02)	NA(0.04)	NA(0.01)	NA(0.04)
μ_θ	0.00(0.01)	-0.01(0.01)	0.00(0.01)	0.00(0.01)	0.00(0.01)
κ	0.02(0.01)	0.02(0.01)	0.02(0.01)	0.02(0.01)	0.02(0.01)

NA: No results are available for these parameters.

Table 4.34: Relative Bias (MSE) for Scenario II, n=50, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.43(0.37)	-0.11(0.75)	-0.31(0.42)	-0.43(0.44)	-0.27(0.41)
β_2	-0.53(0.45)	-0.35(0.66)	-0.43(0.48)	-0.53(0.53)	-0.40(0.50)
β_3	NA(0.25)	NA(0.53)	NA(0.36)	NA(0.18)	NA(0.41)
β_4	NA(0.21)	NA(0.51)	NA(0.32)	NA(0.22)	NA(0.35)
β_5	0.01(0.17)	0.17(0.27)	0.10(0.21)	0.06(0.19)	0.13(0.23)
β_6	0.04(0.25)	0.2(0.45)	0.14(0.32)	0.09(0.32)	0.13(0.36)
β_7	NA(0.08)	NA(0.14)	NA(0.10)	NA(0.08)	NA(0.12)
β_8	NA(0.09)	NA(0.16)	NA(0.11)	NA(0.08)	NA(0.14)
μ_θ	0.04(0.41)	-0.02(0.44)	0.02(0.36)	0.04(0.40)	-0.01(0.40)
κ	0.34(0.18)	0.34(0.19)	0.34(0.19)	0.32(0.17)	0.34(0.19)

NA: No results are available for these parameters.

Table 4.35: Relative Bias (MSE) for Scenario II, n=100, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.14(0.29)	0.15(0.58)	-0.02(0.34)	-0.12(0.43)	0.04(0.36)
β_2	-0.22(0.32)	0.01(0.57)	-0.10(0.37)	-0.22(0.47)	-0.07(0.37)
β_3	NA(0.20)	NA(0.39)	NA(0.28)	NA(0.13)	NA(0.31)
β_4	NA(0.18)	NA(0.30)	NA(0.24)	NA(0.15)	NA(0.26)
β_5	0.05(0.05)	0.19(0.10)	0.12(0.06)	0.10(0.06)	0.14(0.06)
β_6	0.09(0.07)	0.23(0.13)	0.15(0.09)	0.13(0.07)	0.17(0.09)
β_7	NA(0.02)	NA(0.02)	NA(0.02)	NA(0.00)	NA(0.02)
β_8	NA(0.02)	NA(0.02)	NA(0.03)	NA(0.01)	NA(0.03)
μ_θ	0.06(0.10)	-0.01(0.11)	0.03(0.10)	0.06(0.10)	0.01(0.10)
κ	0.09(0.02)	0.10(0.02)	0.10(0.02)	0.09(0.02)	0.10(0.03)

NA: No results are available for these parameters.

Table 4.36: Relative Bias (MSE) for Scenario II, n=250, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.13(0.12)	-0.01(0.16)	-0.07(0.12)	-0.09(0.18)	-0.05(0.13)
β_2	-0.12(0.09)	0.01(0.11)	-0.05(0.09)	-0.06(0.13)	-0.04(0.09)
β_3	NA(0.05)	NA(0.05)	NA(0.06)	NA(0.02)	NA(0.07)
β_4	NA(0.05)	NA(0.05)	NA(0.06)	NA(0.02)	NA(0.06)
β_5	-0.01(0.01)	0.04(0.01)	0.02(0.01)	0.02(0.01)	0.02(0.01)
β_6	0.02(0.02)	0.07(0.02)	0.05(0.02)	0.04(0.02)	0.05(0.02)
β_7	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_8	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
μ_θ	0.04(0.04)	0.01(0.03)	0.02(0.03)	0.04(0.04)	0.02(0.03)
κ	0.01(0.01)	0.01(0.01)	0.01(0.01)	0.01(0.01)	0.01(0.01)

NA: No results are available for these parameters.

Table 4.37: Relative Bias (MSE) for Scenario III, n=50, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-1.01(13.14)	1.63(18.85)	-1.02(17.15)	-0.96(15.78)	-0.86(8.28)
β_2	-0.93(13.64)	-1.53(15.83)	-0.96(12.26)	-1.08(17.08)	-1.06(18.61)
β_3	-0.96(11.28)	0.83(10.29)	-0.94(11.31)	-1.07(11.43)	-1.11(12.02)
β_4	-1.00(13.84)	-1.33(18.97)	-0.93(11.73)	-1.12(14.42)	-0.98(12.43)
β_5	NA(2.83)	NA(2.97)	NA(2.27)	NA(2.36)	NA(1.03)
β_6	NA(3.42)	NA(5.94)	NA(5.92)	NA(5.04)	NA(1.03)
β_7	NA(3.52)	NA(3.31)	NA(4.65)	NA(5.03)	NA(0.91)
β_8	NA(2.84)	NA(3.91)	NA(5.58)	NA(3.88)	NA(1.08)
β_9	-1.03(11.10)	1.82(13.63)	-0.82(10.28)	-0.82(10.72)	-0.86(8.41)
β_{10}	-0.89(10.98)	-1.60(12.93)	-0.92(12.45)	-1.08(13.85)	-0.83(12.27)
β_{11}	-1.04(10.00)	-1.61(11.70)	-1.12(10.97)	-1.06(13.38)	-0.93(9.24)
β_{12}	-0.94(10.22)	-1.47(11.94)	-0.98(10.39)	-0.97(10.14)	-0.89(9.01)
β_{13}	NA(4.75)	NA(4.43)	NA(5.38)	NA(5.01)	NA(1.14)
β_{14}	NA(4.68)	NA(6.14)	NA(7.50)	NA(7.15)	NA(1.01)
β_{15}	NA(5.41)	NA(4.64)	NA(6.16)	NA(10.22)	NA(1.70)
β_{16}	NA(4.68)	NA(4.50)	NA(6.01)	NA(6.61)	NA(1.27)
μ_θ	-0.10(0.12)	-0.10(0.10)	-0.07(0.09)	-0.13(0.12)	-0.09(0.12)
κ	0.09(0.07)	-0.18(0.07)	-0.19(0.07)	-0.18(0.05)	0.27(0.19)

NA: No results are available for these parameters.

Table 4.38: Relative Bias (MSE) for Scenario III, n=100, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.88(10.35)	-1.47(10.15)	-0.84(11.11)	-0.82(11.79)	-0.83(7.11)
β_2	-0.86(9.01)	0.49(7.74)	-0.74(6.96)	-1.02(13.8)	-0.94(11.49)
β_3	-0.89(10.16)	0.72(7.46)	-0.78(9.22)	-1.01(10.98)	-1.05(10.13)
β_4	-0.95(8.54)	-1.05(11.93)	-0.89(8.70)	-1.09(9.84)	-0.96(10.05)
β_5	NA(2.78)	NA(2.13)	NA(2.17)	NA(2.26)	NA(0.94)
β_6	NA(3.14)	NA(4.48)	NA(5.08)	NA(4.53)	NA(0.97)
β_7	NA(2.74)	NA(3.01)	NA(4.25)	NA(4.64)	NA(0.82)
β_8	NA(2.70)	NA(3.89)	NA(5.28)	NA(3.18)	NA(1.48)
β_9	-0.90(10.02)	0.20(10.55)	-0.79(9.52)	-0.73(9.22)	-0.83(7.17)
β_{10}	-0.82(9.73)	1.25(10.33)	-0.89(11.94)	-0.83(13.68)	-0.75(10.68)
β_{11}	-1.00(7.48)	-1.31(10.16)	-1.03(9.86)	-0.98(11.65)	-0.78(7.36)
β_{12}	-0.88(9.95)	-0.17(8.89)	-0.82(9.09)	-0.90(9.35)	-0.84(8.62)
β_{13}	NA(4.04)	NA(4.44)	NA(4.54)	NA(4.93)	NA(1.04)
β_{14}	NA(4.58)	NA(6.02)	NA(6.52)	NA(6.85)	NA(0.86)
β_{15}	NA(4.78)	NA(3.97)	NA(5.65)	NA(9.49)	NA(1.55)
β_{16}	NA(3.24)	NA(3.45)	NA(5.40)	NA(6.15)	NA(1.07)
μ_θ	-0.10(0.08)	-0.10(0.07)	-0.10(0.07)	-0.11(0.11)	-0.03(0.07)
κ	-0.06(0.05)	-0.11(0.05)	-0.09(0.05)	-0.08(0.05)	-0.09(0.04)

NA: No results are available for these parameters.

Table 4.39: Relative Bias (MSE) for Scenario III, n=250, $\kappa = 1$.

Parameters	Methods				
	M_1	M_2	M_3	M_4	M_5
β_1	-0.47(6.22)	-0.18(3.02)	-0.55(9.44)	-0.27(2.53)	-0.45(4.05)
β_2	-0.44(6.98)	-0.19(3.49)	-0.55(4.48)	-0.4(5.05)	-0.49(7.01)
β_3	-0.64(7.65)	-0.18(2.97)	-0.63(8.23)	-0.38(3.80)	-0.50(3.50)
β_4	-0.51(7.91)	-0.14(5.51)	-0.54(7.70)	-0.35(5.56)	-0.51(7.06)
β_5	NA(2.71)	NA(1.13)	NA(2.54)	NA(0.75)	NA(0.48)
β_6	NA(2.14)	NA(0.56)	NA(2.74)	NA(0.92)	NA(0.43)
β_7	NA(2.50)	NA(1.43)	NA(4.06)	NA(1.38)	NA(0.86)
β_8	NA(1.90)	NA(0.63)	NA(2.76)	NA(0.66)	NA(0.50)
β_9	-0.46(8.94)	-0.04(0.73)	-0.28(7.17)	-0.22(1.14)	-0.2(2.35)
β_{10}	-0.36(7.71)	-0.05(1.08)	-0.39(10.61)	-0.24(1.50)	-0.26(2.21)
β_{11}	-0.42(7.37)	-0.08(0.76)	-0.46(9.13)	-0.25(1.10)	-0.20(2.35)
β_{12}	-0.35(6.01)	-0.06(7.12)	-0.45(8.35)	-0.19(1.04)	-0.24(3.03)
β_{13}	NA(3.15)	NA(0.07)	NA(3.29)	NA(0.01)	NA(0.53)
β_{14}	NA(2.17)	NA(0.06)	NA(3.59)	NA(1.02)	NA(1.63)
β_{15}	NA(2.82)	NA(0.04)	NA(3.68)	NA(3.42)	NA(1.23)
β_{16}	NA(2.68)	NA(0.09)	NA(3.58)	NA(0.03)	NA(1.04)
μ_θ	-0.02(0.03)	0.03(0.02)	-0.03(0.03)	0.03(0.02)	0.01(0.02)
κ	-0.01(0.03)	0.12(0.02)	-0.03(0.04)	0.07(0.01)	0.04(0.02)

NA: No results are available for these parameters.

Table 4.40: Relative Bias (MSE) for Scenario IV, n=50, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.83(0.66)	-0.81(0.78)	-0.94(0.68)	-0.74(0.76)	-0.68(0.59)
β_2	-0.73(0.62)	-0.04(0.51)	-0.86(1.19)	-0.89(1.33)	-0.75(0.68)
β_3	-0.89(0.74)	0.92(0.89)	-0.99(1.35)	-0.76(1.33)	-0.76(0.76)
β_4	-0.90(0.74)	-0.05(0.77)	-0.86(1.04)	-0.67(0.72)	-0.67(0.76)
β_5	NA(0.18)	NA(0.81)	NA(0.63)	NA(0.30)	NA(0.39)
β_6	NA(0.17)	NA(0.38)	NA(0.40)	NA(0.22)	NA(0.23)
β_7	NA(0.19)	NA(0.32)	NA(0.60)	NA(0.23)	NA(0.22)
β_8	NA(0.11)	NA(0.38)	NA(0.46)	NA(0.26)	NA(0.16)
β_9	-0.70(0.74)	-0.49(0.61)	-1.09(1.87)	-0.34(0.63)	-0.33(0.37)
β_{10}	-0.76(0.77)	0.91(0.80)	-0.84(1.40)	-0.46(0.73)	-0.35(0.37)
β_{11}	-0.73(0.92)	-0.85(0.95)	-0.99(0.98)	-0.53(0.57)	-0.39(0.48)
β_{12}	-0.90(1.05)	0.70(0.92)	-0.74(1.92)	-0.55(0.58)	-0.47(0.49)
β_{13}	NA(0.20)	NA(0.80)	NA(1.07)	NA(0.83)	NA(0.25)
β_{14}	NA(0.24)	NA(0.66)	NA(0.82)	NA(0.07)	NA(0.15)
β_{15}	NA(0.27)	NA(0.43)	NA(0.83)	NA(0.08)	NA(0.29)
β_{16}	NA(0.31)	NA(0.43)	NA(0.79)	NA(0.06)	NA(0.34)
μ_θ	-0.45(1.11)	-0.83(1.14)	-0.56(1.29)	-0.09(0.61)	-0.17(0.51)
κ	0.10(0.09)	-0.43(0.21)	-0.09(0.11)	0.37(0.31)	0.28(0.25)

NA: No results are available for these parameters.

Table 4.41: Relative Bias (MSE) for Scenario IV, n=100, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.46(0.39)	-0.08(0.72)	-0.44(0.44)	-0.46(0.44)	-0.32(0.33)
β_2	-0.52(0.44)	-0.19(0.51)	-0.43(0.56)	-0.44(0.44)	-0.43(0.38)
β_3	-0.51(0.37)	-0.04(0.40)	-0.47(0.45)	-0.42(0.35)	-0.35(0.35)
β_4	-0.50(0.41)	-0.15(0.48)	-0.39(0.41)	-0.50(0.38)	-0.44(0.44)
β_5	NA(0.21)	NA(0.33)	NA(0.26)	NA(0.10)	NA(0.28)
β_6	NA(0.14)	NA(0.15)	NA(0.17)	NA(0.20)	NA(0.20)
β_7	NA(0.10)	NA(0.31)	NA(0.27)	NA(0.16)	NA(0.21)
β_8	NA(0.11)	NA(0.07)	NA(0.19)	NA(0.13)	NA(0.24)
β_9	-0.18(0.42)	0.18(0.11)	-0.17(0.32)	-0.10(0.12)	-0.04(0.18)
β_{10}	-0.19(0.21)	0.09(0.16)	-0.24(0.60)	-0.07(0.11)	-0.09(0.15)
β_{11}	-0.21(0.23)	0.08(0.09)	-0.28(0.60)	-0.13(0.12)	-0.08(0.12)
β_{12}	-0.24(0.27)	0.22(0.11)	-0.25(0.78)	-0.11(0.11)	-0.06(0.13)
β_{13}	NA(0.05)	NA(0.03)	NA(0.14)	NA(0.01)	NA(0.08)
β_{14}	NA(0.09)	NA(0.03)	NA(0.17)	NA(0.00)	NA(0.08)
β_{15}	NA(0.22)	NA(0.02)	NA(0.34)	NA(0.03)	NA(0.11)
β_{16}	NA(0.21)	NA(0.02)	NA(0.10)	NA(0.01)	NA(0.09)
μ_θ	0.07(0.24)	0.11(0.07)	-0.03(0.40)	0.15(0.19)	0.05(0.17)
κ	0.08(0.05)	0.33(0.12)	0.07(0.08)	0.18(0.05)	0.13(0.06)

NA: No results are available for these parameters.

Table 4.42: Relative Bias (MSE) for Scenario IV, n=250, $\kappa = 1$.

	Methods				
Parameters	M_1	M_2	M_3	M_4	M_5
β_1	-0.20(0.19)	0.03(0.30)	-0.14(0.19)	-0.16(0.26)	-0.12(0.19)
β_2	-0.27(0.20)	-0.10(0.31)	-0.22(0.20)	-0.25(0.29)	-0.21(0.20)
β_3	-0.24(0.19)	-0.02(0.30)	-0.18(0.19)	-0.20(0.30)	-0.16(0.2)
β_4	-0.18(0.18)	0.07(0.37)	-0.11(0.19)	-0.12(0.29)	-0.09(0.20)
β_5	NA(0.11)	NA(0.17)	NA(0.12)	NA(0.08)	NA(0.13)
β_6	NA(0.07)	NA(0.11)	NA(0.08)	NA(0.04)	NA(0.09)
β_7	NA(0.09)	NA(0.13)	NA(0.11)	NA(0.05)	NA(0.12)
β_8	NA(0.09)	NA(0.13)	NA(0.10)	NA(0.07)	NA(0.11)
β_9	-0.03(0.03)	0.11(0.05)	0.01(0.03)	0(0.03)	0.02(0.03)
β_{10}	-0.04(0.04)	0.10(0.06)	0.00(0.04)	-0.01(0.04)	0.01(0.04)
β_{11}	-0.02(0.03)	0.12(0.05)	0.02(0.03)	0.01(0.03)	0.03(0.03)
β_{12}	-0.03(0.02)	0.11(0.04)	0.01(0.03)	0.00(0.03)	0.02(0.03)
β_{13}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_{14}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_{15}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.00)	NA(0.01)
β_{16}	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.01)	NA(0.02)
μ_θ	0.05(0.03)	-0.02(0.03)	0.03(0.03)	0.05(0.03)	0.02(0.03)
κ	0.01(0.01)	0.01(0.01)	0.01(0.01)	0.01(0.01)	0.01(0.01)

NA: No results are available for these parameters.

Table 4.43: Performances in variable selection under Scenario I and III. $\kappa = 3$.

		Scenario I					Scenario III				
n	Measures	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
50	ACI	3.720	3.680	3.730	3.670	3.670	4.455	4.957	4.055	5.000	4.567
	AFE	0.280	0.320	0.270	0.330	0.330	3.545	3.043	3.945	3.000	3.433
	CIR	0.740	0.700	0.740	0.700	0.720	0.000	0.000	0.000	0.000	0.000
100	ACI	3.990	3.980	3.990	3.980	3.990	7.122	6.959	7.163	6.980	7.020
	AFE	0.010	0.020	0.010	0.020	0.010	0.878	1.041	0.837	1.020	0.980
	CIR	0.990	0.980	0.990	0.980	0.990	0.490	0.449	0.510	0.408	0.449
250	ACI	4.000	4.000	4.000	4.000	4.000	8.000	8.000	8.000	8.000	8.000
	AFE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	CIR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4.44: Performances in variable selection under Scenario II and IV. $\kappa = 3$.

		Scenario II					Scenario IV				
n	Measures	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
50	ACI	3.100	3.070	3.190	2.700	3.220	4.850	4.820	4.930	4.530	4.960
	AFE	0.900	0.930	0.810	1.300	0.780	3.150	3.180	3.070	3.470	3.040
	CIR	0.290	0.240	0.340	0.090	0.350	0.000	0.000	0.000	0.000	0.000
100	ACI	3.850	3.830	3.860	3.610	3.870	6.820	6.670	6.840	6.040	6.870
	AFE	0.150	0.170	0.140	0.390	0.130	1.180	1.330	1.160	1.960	1.130
	CIR	0.860	0.840	0.870	0.630	0.880	0.230	0.180	0.250	0.040	0.260
250	ACI	4.000	4.000	4.000	4.000	4.000	8.000	8.000	8.000	8.000	8.000
	AFE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	CIR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4.45: Performances in variable selection under Scenario I and III. $\kappa = 1$.

		Scenario I					Scenario III				
n	Measures	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
50	ACI	0.490	1.000	0.450	0.230	0.424	0.040	0.030	0.025	0.020	0.020
	AFE	3.510	3.000	3.550	3.770	3.576	7.960	7.970	7.975	7.980	7.980
	CIR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
100	ACI	2.200	2.180	2.100	1.950	1.540	0.190	0.150	0.070	0.030	3.220
	AFE	1.800	1.820	1.900	2.050	2.460	7.810	7.850	7.930	7.970	4.780
	CIR	0.120	0.090	0.120	0.090	0.110	0.000	0.000	0.000	0.000	0.000
250	ACI	3.840	3.790	3.830	3.750	3.840	3.290	5.472	3.289	5.393	5.420
	AFE	0.160	0.210	0.170	0.250	0.160	4.710	2.528	4.711	2.607	2.580
	CIR	0.850	0.810	0.840	0.780	0.850	0.020	0.028	0.021	0.036	0.000

Table 4.46: Performances in variable selection under Scenario II and IV. $\kappa = 1$.

		Scenario II					Scenario IV				
n	Measures	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
50	ACI	1.700	1.720	1.740	1.560	1.570	0.510	0.550	0.330	1.357	1.103
	AFE	2.300	2.280	2.260	2.440	2.430	7.490	7.450	7.670	6.643	6.897
	CIR	0.010	0.010	0.010	0.000	0.020	0.000	0.000	0.000	0.000	0.000
100	ACI	2.340	2.360	2.420	2.170	2.420	2.890	4.057	3.010	3.230	3.220
	AFE	1.660	1.640	1.580	1.830	1.580	5.110	3.943	4.990	4.770	4.780
	CIR	0.040	0.020	0.050	0.010	0.050	0.000	0.000	0.000	0.000	0.000
250	ACI	3.390	3.360	3.430	2.890	3.460	5.270	5.250	5.360	4.770	5.420
	AFE	0.610	0.640	0.570	1.110	0.540	2.730	2.750	2.640	3.230	2.580
	CIR	0.470	0.450	0.500	0.170	0.520	0.010	0.000	0.000	0.000	0.000

4.2 Real Datasets Application

In this section, we provide the comprehensive analyses of real-life examples by effectively demonstrating the practical application and effectiveness of the methodologies.

4.2.1 Bayesian Lasso in Linear-Circular Regression Models

In this section we apply our method on two different environmental data sets. The first data set (Section 4.2.1.1) is the air quality data set commonly used in circular literature. The second data set (Section 4.2.1.2) concerns with factors effecting temperature in Sydney where bush fires are regular threat.

Bayesian analyses of these data sets are carried out in OpenBUGS. We consider two different chains with 1,000,000 updates after discarding the first 10,000 iterations as burn-in period. Trace plots and BGR statistic are used for convergence diagnostics and determining the burn-in period. MCMC iterations were run until MC errors based on the Markov chain were less than 5% of the posterior standard deviations. Posterior means (i.e. expectation of the posterior distributions) are used for estimating the model parameters. Finally, 95% equal-tailed posterior CI are employed as a variable selection guide. Note that hyper-hyper priors for r and δ are specified as $Ga(0.01, 0.01)$ in EB-GS method.

In order to explore the performance of Bayesian lasso in terms of prediction accuracy for linear-circular regression models, we use the root mean square errors (RMSEs) of the fitted values. The expressions in Equation 4.5 can be used to calculate RMSEs

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{4.5}$$

where, n is the number of observations, y_i and $\hat{y}_i$ are the actual and predicted response values for i^{th} observation, respectively. Note that only significant covariates are used to obtain predicted response values.

4.2.1.1 Air Quality Index Data Set

Herein we use the air quality index (AQI) data set to assess the performance of Bayesian lasso in a real data application. De Wiest and Della Fiorentina (1975) introduced a method to derive air quality index and constructed a data set consisting of their air quality index and environmental variables such as temperature, wind direc-

tion and speed. Johnson and Wehrly (1978) used a regression model based on temperature and wind direction to predict the AQI based on their data set. Here we consider a regression model with the AQI as the linear dependent variable, temperature (TEMP) and wind speed (WS) as linear independent variables and wind direction (WD) as circular independent variable and apply Bayesian lasso for variable selection. We tested the plausibility of normality assumption for the response data. Shapiro–Wilk goodness of fit test (Shapiro and Wilk, 1965) and its p-value are 0.95 and 0.54 (at significance level 0.05), respectively, implying that normal distribution seems to be a plausible distribution for the data.

Table 4.47 presents results of our Bayesian lasso for AQI analysis. Bayesian lasso with non-informative gamma (Ga(0.1, 0.1)) and EB-GS hyper-prior for the tuning parameter tend to select one covariate, which is wind direction, based on 95% CIs. On the other hand, Bayesian lasso with informative hyper-prior (HC(0,1)) for the tuning parameter tends to select neither one. For all hyper-priors, sign of parameter estimates are consistent with the previous analysis of this data set in the literature. Magnitude of parameter estimates are very close to those in the previous analysis of this data set for Ga(0.1,0.1) and EB-GS hyper-prior. In addition, MC errors and posterior standard deviation (given in parenthesis) measuring posterior uncertainty are generally lower for EB-GS hyper-prior than the others, especially for the tuning parameter. Note that the most important contribution of this data analysis is to reveal the uncertainty about parameter estimation by using posterior standard deviation. When prediction accuracy is concerned, RMSE of Bayesian lasso with EB-GS hyper-prior is notably smaller than the others implying good prediction faculties. In Figure 4.1, the x-axis shows predicted values for AQI, while the y-axis shows the dataset's actual values. The agreement between the values can be checked with the diagonal line in the center of the plot. Because most of the data points are quite close to the diagonal line for Bayesian lasso with non-informative gamma and EB-GS hyper-prior for the tuning parameter, we may conclude that the regression model fits the data reasonably well. This is not fully valid when Bayesian lasso with informative hyper-prior is used because most of the points are far from the diagonal line.

Table 4.47: Results for the AQI data set.

	Ga(0.1,0.1)			HC(0,1)			Ga($\hat{r}$, $\hat{\delta}$)		
Parameters	Estimate	MC Error	95% CI	Par. Est.	MC Error	95% CI	Par. Est.	MC Error	95% CI
Intercept	0.331(0.171)	$5*10^{-4}$	(-0.014, 0.660)	0.363(0.176)	$2*10^{-3}$	(-0.001, 0.686)	0.328(0.160)	$1*10^{-4}$	(0.006,0.642)
TEMP	0.028(0.023)	$7*10^{-5}$	(-0.016, 0.076)	0.024(0.024)	$2*10^{-4}$	(-0.016,0.074)	0.028(0.022)	$2*10^{-5}$	(-0.015, 0.072)
WS	-0.001(0.004)	$8*10^{-6}$	(-0.009, 0.006)	-0.001(0.004)	$1*10^{-5}$	(-0.008,0.007)	-0.001(0.003)	$2*10^{-6}$	(-0.008,0.006)
COS WD	-0.104(0.081)	$1*10^{-4}$	(-0.273, 0.038)	-0.072(0.081)	$9*10^{-4}$	(-0.254, 0.041)	-0.137(0.078)	$1*10^{-5}$	(-0.294,0.014)
SIN WD	0.164(0.081)	$1*10^{-4}$	(0.004, 0.321)	0.117(0.095)	$2*10^{-3}$	(-0.016, 0.304)	0.197(0.072)	$1*10^{-5}$	(0.052, 0.338)
σ^2	0.190(0.021)	$5*10^{-5}$	(0.020, 0.102)	0.057(0.030)	$3*10^{-4}$	(0.021, 0.132)	0.036(0.039)	$2*10^{-5}$	(0.017, 0.081)
λ	2.190(1.136)	$3*10^{-3}$	(0.572, 4.924)	14.720(47.200)	$18*10^{-1}$	(0.699, 116.400)	1.037(0.154)	$3*10^{-5}$	(0.747, 1.352)
RMSE		0.172			0.246			0.169	

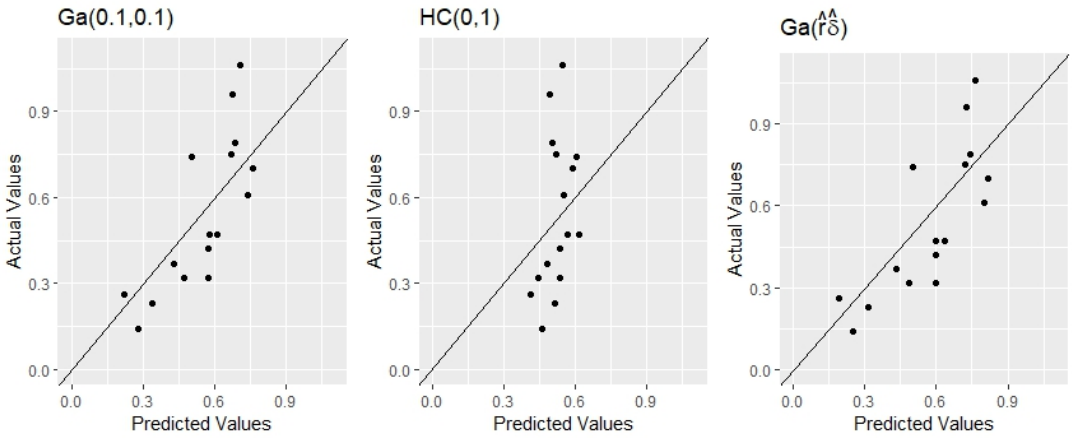

Figure 4.1: Actual vs. Fitted values for the AQI data set.

4.2.1.2 Daily Weather Data Set

The second data set has been retrieved from the data repository of Australian Bureau of Meteorology (BOM). The data set consists of daily weather observations recorded in Sydney, New South Wales between 1 November 2021 and 28 February 2022. This period is a bush fire season in New South Wales according to BOM and thus understanding the fire weather variables help fire prevention and control.

Herein our aim is to illustrate the use of Bayesian lasso in linear-circular regression model for determining covariates that can be used to predict the maximum temperature which is linear response variable. Among the covariates, three of them are linear, bright sunshine hour in the 24 hours to midnight (SUN), fraction of sky obscured by cloud at 9 am (CLOUD9AM) and at 3 pm (CLOUD3PM). The others are circular, direction of strongest gust to midnight (DIRSG) and wind direction averaged over 10 minutes prior to 3 pm (DIRAV) in the 24 hours. Shapiro–Wilk goodness of fit test (Shapiro and Wilk, 1965) and its p-value are 0.98 and 0.16 (at significance level 0.05), respectively, implying that normal distribution seems to be a plausible distribution for the data.

We again perform an extensive comparison of our proposed method with a variety of hyper-priors for the tuning parameter. Hyper-priors for the tuning parameter are the same as in Section 4.2.1.1. Extensive results are presented in Table 4.48. A first clear conclusion is that the behaviors of Bayesian lasso with different hyper-priors for

the tuning parameter are similar to each other, suggesting strength in the inclusion of DIRAV, SUN and CLOUD9AM covariates. Another conclusion is that the Bayesian lasso with EB-GS hyper-prior distribution for λ performs really well providing RMSE that is comparable or smaller than the others. Note that this is a situation where prior information about the model parameters is not available. In this case EB-GS prior is more practical.

Table 4.48: Results for the daily weather data set.

	Ga(0.1,0.1)			HC(0,1)			Ga($\hat{r}$, $\hat{\delta}$)		
Parameters	Par. Est.	MC Error	95% CI	Par. Est.	MC Error	95% CI	Par. Est.	MC Error	95% CI
Intercept	17.970(2.015)	$1*10^{-2}$	(14.000, 21.930)	18.170(2.041)	$1*10^{-2}$	(14.170, 22.190)	18.300(2.046)	$1*10^{-2}$	(14.270, 22.330)
COS DIRAV	2.071(0.606)	$1*10^{-3}$	(0.868, 3.256)	1.971(0.623)	$2*10^{-3}$	(0.7156, 3.175)	1.555(0.619)	$1*10^{-3}$	(0.136, 2.974)
SIN DIRAV	-0.246(0.451)	$1*10^{-3}$	(-1.179, 0.616)	-0.240(0.431)	$9*10^{-4}$	(-1.145, 0.5774)	-0.235(0.411)	$9*10^{-4}$	(-1.110, 0.536)
COS DIRSG	0.183(0.512)	$1*10^{-3}$	(-0.815, 1.256)	0.204(0.494)	$1*10^{-3}$	(-0.752, 1.250)	0.218(0.473)	$1*10^{-3}$	(-0.683, 1.242)
SIN DIRSG	-0.666(0.474)	$1*10^{-3}$	(-1.627, 0.198)	-0.614(0.464)	$1*10^{-3}$	(-1.574, 0.203)	-0.575(0.452)	$1*10^{-3}$	(-1.524, 0.206)
SUN	0.495(0.145)	$8*10^{-4}$	(0.211, 0.779)	0.482(0.146)	$8*10^{-4}$	(0.194, 0.769)	0.474(0.145)	$8*10^{-4}$	(0.185, 0.762)
CLOUD9AM	0.514(0.177)	$7*10^{-4}$	(0.168, 0.863)	0.491(0.181)	$8*10^{-4}$	(0.133, 0.846)	0.475(0.174)	$8*10^{-4}$	(0.109, 0.833)
CLOUD3PM	-0.031(0.169)	$7*10^{-4}$	(-0.369, 0.304)	-0.034(0.167)	$7*10^{-4}$	(-0.368, 0.295)	-0.036(0.166)	$6*10^{-4}$	(-0.369, 0.294)
σ^2	7.415(1.180)	$2*10^{-3}$	(5.456, 10.060)	7.545(1.224)	$3*10^{-3}$	(5.522, 10.300)	7.660(1.162)	$2*10^{-3}$	(5.581, 10.500)
λ	3.017(1.056)	$2*10^{-3}$	(1.281, 5.370)	3.940(1.888)	$1*10^{-2}$	(1.406, 8.549)	4.743(1.537)	$1*10^{-3}$	(1.639, 10.480)
RMSE		2.701			2.740			2.695	

CHAPTER 5

CONCLUSION

Selection of an appropriate approximating model containing the most relevant variables is critical to statistical inference. Current literature is abundant with variable selection methods for regression models with linear variables, but when modeling with circular variables is of concern, these usual variable selection methods employed for linear data may not be appropriate, situation similar to the other usual statistical tools not being suitable for analysis of circular data.

The aim of this thesis was to initiate a work on Bayesian penalization techniques such as the Bayesian lasso and the Bayesian adaptive lasso for variable selection in circular regression models to reduce model complexity and improve the prediction ability of the model. In this thesis, within the framework of linear-circular regression models, we propose an adaptation of the Bayesian lasso for this purpose and develop a novel hyper-hyper parameter elicitation for the tuning parameter, when a gamma prior is used as hyper-prior distribution. Within the framework of circular-(circular, linear) regression models, we propose an adaptation of the Bayesian lasso and the Bayesian adaptive lasso. We also develop some alternative Bayesian adaptive lasso methods. Our methods allow for individual shrinking coefficients for each coefficient. Based on simulations and empirical results, we highlight the following attractive properties and some future works.

- In the framework of linear-circular regression models;
  - The adaptation of Bayesian lasso can straightforwardly be implemented by using a simple Gibbs sampling procedure for linear-circular regression models.

- Our adaption with EB-GS hyper-prior distribution for λ performs satisfactorily as variable selection method identifying parsimonious models. It select parsimonious models of appropriate dimension by excluding the insignificant covariates and including the significant covariates.
- Our method seems to provide a clearer distinction between significant and insignificant covariates than the Bayesian lasso with usual hyper-prior distribution for λ.
- Another important characteristic of our proposed method is that it is sufficient to give just non-informative hyper-hyper prior distributions for hyper-hyper parameters, i.e. r and δ.
- Real life examples show that our proposed method has smaller RMSEs and MC errors than conventional hyper-prior distributions.
- Results demonstrate that our proposed method has the better performance in terms of variable selection, parameter estimations and prediction accuracy than non-informative gamma and informative HC distributions.
- Our adapted Bayesian lasso method for linear-circular regression models works efficiently and the EB-GS hyper-prior is promising for choosing hyper-hyper parameter for the λ.
- The proposed EB-GS hyper-prior is general and can be used to select hyper-hyper parameters in other hyper-prior such as HC prior.

- In the framework of circular-(circular, linear) regression models;
  - The adaptation of Bayesian lasso and Bayesian adaptive lasso can straightforwardly be implemented by using a simple Gibbs sampling procedure for circular-(circular, linear) regression models.
  - The proposed alternatives of Bayesian adaptive lasso can straightforwardly be implemented by using a simple Gibbs sampling procedure for circular-(circular, linear) regression models.
  - The proposed alternatives of Bayesian adaptive lasso can effectively identify and estimate significant variables by assigning different shrinkage levels to different coefficients.

  - They provide better flexibility in capturing the underlying associations of covariates with the response.
  - The adapted and proposed methods are general and can be employed in Euclidean and circular regression settings.

  To conclude, results show that both the adapted and the proposed methods perform very well in terms of prediction accuracy and variable selection in circular regression models.

- The following cases can be considered as future work;

  - when the underlying distribution of the circular dependent variable is asymmetric and or multimodal,
  - cases where the number of independent variables, both linear and circular, is much larger when there is curse of high dimensionality as in small small n-big p,
  - when sample size is small,
  - when there is correlation or grouping between independent variables.

  Assessing these situations will reveal both the limitations of the proposed and adapted methods and how they work in different situations. This will lead to good guidance for the users of these methods.

REFERENCES

Abuzaid, A.H. and Allahham, N.R. (2015). Simple circular regression model assuming wrapped Cauchy error. *Pakistan Journal of Statistics*, 31(4), 385-398

Andrews, D. F., and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society Series B-methodological*, 36(1), 99–102.

Bae, K., and Mallick, B. K. (2004). Gene selection using a two-level hierarchical Bayesian model. *Bioinformatics*, 20(18), 3423-3430.

Bhattacharya, S., and Sengupta, A. (2009). Bayesian analysis of semiparametric linear-circular models. *Journal of agricultural, biological, and environmental statistics*, 14, 33-65.

Bühlmann, P., and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications.* Springer Science and Business Media.

Brooks, S., and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455.

Camli, O., and Kalaylioglu, Z. (2021). Bayesian predictive model selection in circular random effects models with applications in ecological and environmental studies. *Environmental and Ecological Statistics*, 28(1), 21–34.

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.

Chen, X., Wang, Z. J., and McKeown, M. J. (2011). A Bayesian Lasso via reversible-jump MCMC. *Signal Processing*, 91(8), 1920-1932.

De Wiest, F., and Della Fiorentina, H. (1975). Suggestions for a Realistic definition of an air quality index relative to hydro- carbonaceous matter associated with airborne particles. *Atmospheric Environment*, 9(10), 951-954.

D'elia, A. (2001). A statistical model for orientation mechanism. *Statistical Methods and Applications*, 10, 157-174.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2013). Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40(2), 238-255.

Downs, T. D. and Mardia, K. V. (2002). Circular regression. *Biometrika*, 89(3), 683–698.

Efron, B., Hastie, T., Johnstone, I. M., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32(2).

Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348–1360.

Fan, Y., and Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society Series B-statistical Methodology*, 75(3), 531–552.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 209-230.

Fernández-Durán, J. J., and Gregorio-Domínguez, M. M. (2014). Modeling angles in proteins and circular genomes using multivariate angular distributions based on multiple nonnegative trigonometric sums. *Statistical Applications in Genetics and Molecular Biology*, 13(1), 1–18.

Figueiredo, M. A. (2003). Adaptive sparseness for supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 25(9), 1150-1159.

Fisher, N. I., and Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, 48(3), 665.

Fisher, N.I. (1993). *Statistical analysis circular data.* Cambridge, England, Cambridge University Press.

Fisher, N. I., and Lee, A. (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(2), 327-339.

Flynn, C. J., Hurvich, C. M., and Simonoff, J. S. (2013). Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models. *Journal of the American Statistical Association*, 108(503), 1031–1043.

Geisser, S., and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365), 153–160.

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

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515–534.

Goudie, R. J. B., Turner, R. M., De Angelis, D., and Thomas, A. (2020) MultiBUGS: A parallel implementation of the BUGS modelling framework for faster Bayesian inference. *Journal of Statistical Software*, 95(7).

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

Hall, P., Lee, E. R., and Park, B. U. (2009). Bootstrap-based penalty choice for the lasso achieving oracle performance. *Statistica Sinica*, 19(2), 449–471.

Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4), 835–845.

Hernandez-Stumpfhauser, D., Breidt, J.F., Opsomer, J.D. (2016). Hierarchical Bayesian Small Area Estimation for Circular Data. Colorado State University.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.

Hrushesky, W. J. M. (1985). Circadian timing of cancer chemotherapy. *Science*, 228(4695), 73–75.

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

Inoue, T., and Nagata, Y. (2016). Study of model evaluation method and of selection method of tuning parameter in Lasso. *Total Quality Science*, 2(1), 27–35.

Ishwaran, H., and James, L. F. (2002). Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical statistics*, 11(3), 508-532.

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

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning: with Applications in R*. New York. Springer Science and Business Media.

Jha, J., and Biswas, A. (2017). Multiple circular-circular regression. *Statistical Modelling*, 17(3), 142-171.

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

Jones, M. C., and Pewsey, A. (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, 100(472), 1422–1428.

Jones, M.C., and Silverman, B.W. (1989). An orthogonal series density estimation approach to reconstructing positron emission tomography images. *Journal of Applied Statistics*, 16(2), 177–191.

Kato, S., Shimizu, K. and Shieh, G. (2008). A circular-circular regression model. *Statistica Sinica*, 18(2), 633-645.

Kawano, S., Hoshina, I., Shimamura, K., and Konishi, S. (2015). Predictive model selection criteria for Bayesian lasso regression. *Journal of the Japanese Society of Computational Statistics*, 28(1), 67-82.

Kim, S., and SenGupta, A. (2015). Inverse circular-linear/linear-circular regression. *Communications in Statistics*, 44(22), 4772–4782.

Kim, S., and SenGupta, A. (2017). Multivariate multiple circular regression. *Journal of Statistical Computation and Simulation*, 87(7), 1277-1291.

Kim, S. (2009). *Inverse circular regression with possibly asymmetric circular error distribution*. University of California, Riverside.

Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, 5(2), 369-411.

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

Leng, C., Tran, M., and Nott, D. J. (2013). Bayesian adaptive Lasso. *Annals of the Institute of Statistical Mathematics*, 66(2), 221–244.

Lowrey, P. L., Shimomura, K., Antoch, M. P., Yamazaki, S., Zemenides, P. D., Ralph, M. R., Menaker, M., and Takahashi, J. S. (2000). Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science*, 288(5465), 483–491.

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

Lykou, A., and Ntzoufras, I. (2013). On Bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing*, 23(3), 361–390.

Mallick, H., and Yi, N. (2014). A new Bayesian lasso. *Statistics and Its Interface*, 7(4), 571–582.

Mardia, K. V. (1972). *Statistics of directional data*. New York, Academic Press.

Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, 63(2), 403-405.

Mardia, K. V., Hughes, G., Taylor, C., and Singh, H. (2008). A multivariate von Mises distribution with applications to bioinformatics. *Canadian Journal of Statistics*, 36(1), 99–109.

Mardia, K.V., and Jupp, P.E. (1999). *Directional statistics*. New York, John Wiley and Sons.

Maruotti, A., Punzo, A., Mastrantonio, G., and Lagona, F. (2015). 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(6), 1725–1740.

Maruotti, A. (2016). Analyzing longitudinal circular data by projected normal models: a semi-parametric approach based on finite mixture models. *Environmental and Ecological Statistics*, 23(2), 257–277.

McMillan, G. P., Hanson, T., Saunders, G. H., and Gallun, F. J. (2013). A two-component circular regression model for repeated measures auditory localization data. *Applied Statistics*, 62(4), 515–534.

Mulder, J., and Pericchi, L. R. (2018). The matrix-F prior for estimating and testing covariance matrices. *Bayesian Analysis*, 13(4), 1189–1210.

Murray, W., Gill, P. and Wright, M. (1981). *Practical optimization*. New York: Academic.

Nuñez-Antonio, G., and Gutiérrez-Peña, E. (2005). A Bayesian analysis of directional data using the projected normal distribution. *Journal of Applied Statistics*, 32(10), 995-1001.

Nuñez-Antonio, G., Gutiérrez-Peña, E., and Escarela, G. (2011). A Bayesian regression model for circular data based on the projected normal distribution. *Statistical Modelling*, 11(3), 185-201.

Park, T., and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681–686.

Polson, N. G., and Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4).

Presnell, B., Morrison, S. P., and Littell, R. C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, 93(443), 1068-1077.

Ravindran, P., and Ghosh, S. K. (2011). Bayesian analysis of circular data using wrapped distributions. *Journal of Statistical Theory and Practice*, 5(4), 547-561.

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

Rueda, C., Fernández, M. A., Barragán, S., Mardia, K. V., and Peddada, S. D. (2016). Circular piecewise regression with applications to cell-cycle data. *Biometrics*, 72(4), 1266–1274.

Sarma, Y., and Jammalamadaka, S. (1993). Circular regression. *Statistical Science and Data Analysis*, 34, 109-128.

SenGupta, A. (2004). On the constructions of probability distributions for directional data. *Bulletin of the Indian Mathematical Society*, 96(2), 139-154.

SenGupta, A., and Ugwuowo, F. I. (2006). Asymmetric circular-linear multivariate regression models with applications to environmental data. *Environmental and Ecological Statistics*, 13(3), 299–309.

SenGupta, A., Kim, S., and Arnold, B. C. (2013). Inverse circular-circular regression. *Journal of Multivariate Analysis*, 119, 200–208.

SenGupta, A., and Kim, S. (2016). Statistical inference for homologous gene pairs between two circular genomes: a new circular–circular regression model. *Statistical Methods and Applications*, 25, 421-432.

SenGupta, A., and Bhattacharya, S. (2015). Finite mixture-based Bayesian analysis of linear-circular models. *Environmental and Ecological Statistics*, 22(4), 667–679.

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591-611.

Shearman, L. P., Sathyanarayanan, S., Weaver, D. R., Maywood, E. S., Chaves, I., Zheng, B., Kume, K., Lee, C., Van Der, G. T. J., Horst, Hastings, M. H., and Reppert, S. M. (2000). Interacting molecular loops in the mammalian circadian clock. *Science*, 288(5468), 1013–1019.

Sikaroudi, A. E., and Park, C. (2019). A mixture of linear-linear regression models for a linear-circular regression. *Statistical Modelling*, 21(3), 220–243.

Snelson, E., Ghahramani, Z., and Rasmussen, C. (2003). Warped gaussian processes. *Advances in neural information processing systems*, 16.

Tateishi, S., Matsui, H., and Konishi, S. (2010). Nonlinear regression modeling via the lasso-type regularization. *Journal of Statistical Planning and Inference*, 140(5), 1125–1134.

Team, R. D. C. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B-methodological*, 58(1), 267–288.

Van Erp, S., Oberski, D. L., and Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, 89, 31–50.

Wang, F., and Gelfand, A. E. (2013). Directional data analysis under the general projected normal distribution. Statistical methodology, 10(1), 113-127.

Wang, H., Li, R., and Tsai, C. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, 94(3), 553–568.

Wang, H., Li, B., and Leng, C. (2008). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society Series B-statistical Methodology*, 71(3), 671–683.

Yu, Y., and Feng, Y. (2014). Modified cross-validation for penalized high-dimensional linear regression models. *Journal of Computational and Graphical Statistics*, 23(4), 1009-1027.

Yuan, M., and Lin, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100(472), 1215-1225.

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301-320.

APPENDIX A

DERIVATION OF FULL CONDITIONAL DISTRIBUTIONS

This appendix delves into the process of deriving full conditional distributions.

A.1 Bayesian Lasso in Linear-Circular Regression Models

Assume $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)^T$ is a $n \times 1$ vector of linear response. Let $\Omega_1 = \{\beta_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}^2, \lambda^2\}$ be parameter space, where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_{2p})^T$ is $(2p) \times 1$ vector of regression coefficients, and $\boldsymbol{\tau}^2 = (\tau_1^2, \tau_2^2 ..., \tau_{2p}^2)$, with p denoting the total number of circular covariates. Then the joint probability density function (pdf) is given by:

$$\begin{aligned} p(\boldsymbol{Y}, \Omega_1) =& L(\beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{Y}) \cdot p(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \sigma^2) \cdot p(\sigma^2 \mid \alpha_\sigma, \beta_\sigma) \cdot p(\boldsymbol{\tau}^2 \mid \lambda^2) \cdot \\ & p(\lambda^2 \mid r, \delta) \cdot p(r) \cdot p(\delta) \cdot p(\beta_0) \end{aligned}$$

Here's the detailed breakdown of each component:

- **Likelihood**:

$$\begin{aligned} L(\beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{Y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2\sigma^2}(\mathbf{y} - \beta_0\mathbf{1} - \boldsymbol{X}\boldsymbol{\beta})^T(\mathbf{y} - \beta_0\mathbf{1} - \boldsymbol{X}\boldsymbol{\beta})), \\ -\infty < \beta_0 < \infty, -\infty < \boldsymbol{\beta} < \infty, \sigma^2 > 0, \end{aligned}$$

  where $\mathbf{1}$ is a vector of ones with the same length as $\boldsymbol{Y}$, and $\boldsymbol{X}$ is $n \times (2p)$ matrix of circular covarites defined in Section 3.3.1.

- **Prior for $\boldsymbol{\beta}$**:

$$p(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \sigma^2) = \frac{1}{(2\pi)^{2p/2} \mid \sigma^2 \mathbf{D}_\tau \mid^{1/2}} e^{-\frac{1}{2\sigma^2}\boldsymbol{\beta}^T {\mathbf{D}_\tau}^{-1} \boldsymbol{\beta}}, -\infty \leq \boldsymbol{\beta} < \infty,$$

where $\mathbf{D}_\tau = diag(\tau_1^2, ..., \tau_{2p}^2)$, $|\sigma^2\mathbf{D}_\tau|$ and $(\sigma^2\mathbf{D}_\tau)^{-1}$ are the determinant and the inverse of the covariance matrix $\sigma^2\mathbf{D}_\tau$, respectively.

- **Prior Distribution for σ^2**:

  The inverse gamma distribution with shape parameter α_σ and scale parameter β_σ is given by:

$$p(\sigma^2 \mid \alpha_\sigma, \beta_\sigma) = \frac{\beta_\sigma^a}{\Gamma(\alpha_\sigma)}(\sigma^2)^{-(\alpha_\sigma+1)} \exp\left(-\frac{\beta_\sigma}{\sigma^2}\right), \sigma^2 > 0, \alpha_\sigma > 0, \beta_\sigma > 0,$$

  where $\Gamma(\cdot)$ is the gamma function.

- **Priors for τ_j^2**: For each $j = 1, \ldots, 2p$:

$$p(\tau_j^2 \mid \lambda^2) = \frac{\lambda^2}{2} e^{-\frac{\lambda^2}{2}\tau_j^2}, \tau_j^2 > 0, \lambda^2 > 0,$$

  where λ^2 is rate parameter.

- **Prior for λ^2**:

$$p(\lambda^2|r, \delta) = \frac{\delta^r}{\Gamma(r)}(\lambda^2)^{r-1}e^{-\delta\lambda^2}, \lambda^2 > 0, r > 0, \delta > 0,$$

  where r and δ are shape and rate parameters, respectively.

- **Prior for r**:

$$p(r \mid \alpha_r, \beta_r) = \frac{\alpha_r^{\beta_r}}{\Gamma(\alpha_r)}(r)^{\alpha_r-1}e^{-\beta_r r}, r > 0, \alpha_r > 0, \beta_r > 0,$$

  where α_r and β_r are shape and rate parameters, respectively.

- **Prior for δ**:

$$p(\delta \mid \alpha_\delta, \beta_\delta) = \frac{\alpha_\delta^{\beta_\delta}}{\Gamma(\alpha_\delta)}(\delta)^{\alpha_\delta-1}e^{-\beta_\delta \delta}, \delta > 0, \alpha_\delta > 0, \beta_\delta > 0,$$

  where α_r and β_r are shape and rate parameters, respectively.

- **Prior for β_0**:

  An independent flat prior can be defined for β_0.

The full conditional distribution for each parameter are as follow:

- **Full conditional distribution for $\boldsymbol{\beta}$**:

$$
\begin{aligned}
p(\boldsymbol{\beta} \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\tau}^2, \sigma^2, \lambda^2) &\propto L(\beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{Y}) \cdot p(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \sigma^2) \\
&\propto \exp(-\frac{1}{2\sigma^2}(\tilde{\mathbf{y}} - \boldsymbol{X}\boldsymbol{\beta})^T(\tilde{\mathbf{y}} - \boldsymbol{X}\boldsymbol{\beta})) \cdot \exp(-\frac{1}{2\sigma^2}\boldsymbol{\beta}^T D\tau^{-1}\boldsymbol{\beta}) \\
&\propto \exp(-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^T(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{\beta} - 2\tilde{\mathbf{y}}\boldsymbol{X}\boldsymbol{\beta} + \tilde{\mathbf{y}}^T\tilde{\mathbf{y}} + \boldsymbol{\beta}^T {\boldsymbol{D}_{\boldsymbol{\tau}}}^{-1}\boldsymbol{\beta}) \\
&\propto \exp(-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^T(\boldsymbol{X}^T\boldsymbol{X} + {\boldsymbol{D}_{\boldsymbol{\tau}}}^{-1})\boldsymbol{\beta} - 2\tilde{\mathbf{y}}\boldsymbol{X}\boldsymbol{\beta} + \tilde{\mathbf{y}}^T\tilde{\mathbf{y}}) \\
&\propto \exp(-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^T\boldsymbol{A}\boldsymbol{\beta} - 2\tilde{\mathbf{y}}\boldsymbol{X}\boldsymbol{\beta} + \tilde{\mathbf{y}}^T\tilde{\mathbf{y}}) \\
&\propto \exp(-\frac{1}{2\sigma^2}((\boldsymbol{\beta} - \boldsymbol{A}^{-1}\boldsymbol{X}^T\tilde{\mathbf{y}})^T\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{A}^{-1}\boldsymbol{X}^T\tilde{\mathbf{y}}) \\
&\qquad + \tilde{\mathbf{y}}^T(\boldsymbol{I_n} - \boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^T)\tilde{\mathbf{y}}) \\
&\propto \exp(-\frac{1}{2\sigma^2}((\boldsymbol{\beta} - \boldsymbol{A}^{-1}\boldsymbol{X}^T\tilde{\mathbf{y}})^T\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{A}^{-1}\boldsymbol{X}^T\tilde{\mathbf{y}}),
\end{aligned}
$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \beta_0\mathbf{1}$ and $\boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{X} + {\boldsymbol{D}_{\boldsymbol{\tau}}}^{-1}$. We used the following square to obtain last line of derivation

$$
(\boldsymbol{\beta} - \boldsymbol{A}^{-1}\boldsymbol{X}^T\tilde{\mathbf{y}})^T\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{A}^{-1}\boldsymbol{X}^T\tilde{\mathbf{y}}) = \boldsymbol{\beta}\boldsymbol{A}^T\boldsymbol{\beta} - 2\tilde{\mathbf{y}}\boldsymbol{X}\boldsymbol{\beta} + \tilde{\mathbf{y}}^T(\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{X}^T)\tilde{\mathbf{y}}.
$$

Then,

$$
\boldsymbol{\beta} \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\tau}^2, \sigma^2, \lambda^2 \sim N_{2p}(\boldsymbol{A}^{-1}\boldsymbol{X}^T\tilde{\mathbf{y}}, \sigma^2\boldsymbol{A}^{-1}).
$$

- **Full conditional distribution for σ^2**:

$$
\begin{aligned}
p(\sigma^2 \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \lambda^2) &\propto L(\beta_0, \boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{Y}) \cdot p(\sigma^2 \mid \alpha_\sigma, \beta_\sigma) \\
&\propto (\sigma^2)^{(-(n-1)/2 - 2p/2 - \alpha_\sigma - 1)} \cdot \exp(-\frac{1}{2\sigma^2}(\tilde{\mathbf{y}} - \boldsymbol{X}\boldsymbol{\beta})^T(\tilde{\mathbf{y}} - \boldsymbol{X}\boldsymbol{\beta})) \\
&\quad \cdot \exp(-\frac{1}{2\sigma^2}\boldsymbol{\beta}^T {\boldsymbol{D}_{\boldsymbol{\tau}}}^{-1}\boldsymbol{\beta} + \frac{\beta_\sigma}{\sigma^2}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\sigma^2 \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \lambda^2 \sim\ & Inverse - Gamma((n-1)/2 + p, \\
& ((\boldsymbol{y} - \beta_0 - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \beta_0 - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\beta}^T {\boldsymbol{D}_{\boldsymbol{\tau}}}^{-1}\boldsymbol{\beta})/2 + \beta_\sigma).
\end{aligned}
$$

- **Full conditional distribution for τ_j^2, $j = 1, \ldots, 2p$:**

  The full conditional distribution for τ_j^2 given all other parameters is given by:

$$\begin{aligned} p(\tau_j^2 \mid \boldsymbol{Y}, \beta_0, \beta_j, \tau_{-j}^2, \sigma^2, \lambda^2) &\propto p(\beta_j \mid \tau_j^2, \sigma^2) \cdot p(\tau_j^2 \mid \lambda^2) \\ &\propto (\tau_j^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2/\sigma^2}{\tau_j^2} + \lambda^2 \tau_j^2)} \\ &\propto (1/\tau_j^2)^{-\frac{3}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\sigma^2 \tau_j^2} + \frac{\lambda^2}{1/\tau_j^2})} \\ &\propto (1/\tau_j^2)^{-\frac{3}{2}} \cdot e^{-\frac{\beta_j^2(\frac{1}{\tau_j^2} - \sqrt{\lambda^2\sigma^2/\beta_j^2})^2}{2\sigma^2(1/\tau_j^2)}}. \end{aligned}$$

  This is in the form of the reciprocal of the Inverse-Gaussian distribution. Then

$$\frac{1}{\tau_j^2} \mid \boldsymbol{Y}, \beta_0, \beta_j, \tau_{-j}^2, \sigma^2, \lambda^2 \sim Inverse - Gaussian(\sqrt{\frac{\lambda^2\sigma^2}{\beta_j^2}}, \lambda^2).$$

- **Full conditional distribution for λ^2:**

  The full conditional distribution for λ^2 given all other parameters is given by:

$$\begin{aligned} p(\lambda^2 \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) &\propto p(\tau_1^2, ..., \tau_{2p}^2 \mid \lambda^2) \cdot p(\lambda^2 \mid r, \delta) \\ &\propto \prod_{j=1}^{2p} \frac{\lambda^2}{2} \cdot e^{-\frac{\lambda^2 \tau_j^2}{2}} (\lambda^2)^{r-1} e^{-\delta\lambda^2} \\ &\propto (\lambda^2)^{2p+r-1} \cdot e^{-\lambda^2(\frac{1}{2}\sum_{j=1}^{2p} \tau_j^2 + \delta)}. \end{aligned}$$

  We can recognize this as a gamma density for λ^2:

$$\lambda^2 \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2 \sim Gamma(2p + r, \frac{1}{2}\sum_{j=1}^{m} \tau_j^2 + \delta).$$

- **Full conditional distribution for β_0:**

$$\beta_0 \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}^2, \lambda^2 \sim N(\bar{y}, \sigma^2/n),$$

- **Full conditional distribution for r:**

$$\begin{aligned} p(r \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2, \lambda^2, \delta) &\propto p(\lambda^2 \mid r, \delta) \cdot p(r) \\ &\propto \frac{\delta^r}{\Gamma(r)} (\lambda^2)^{r-1} \cdot r^{\alpha_r - 1} \cdot \exp(-\beta_r r). \end{aligned}$$

- **Full conditional distribution for** δ:

$$\begin{aligned} p(\delta \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2, \lambda^2, r) &\propto p(\lambda^2 \mid r, \delta) \cdot p(\delta) \\ &\propto \delta^r \exp(-\delta\lambda^2)\delta^{\alpha_\delta - 1} \cdot \exp(-\beta_\delta \delta) \\ &\propto \delta^{(r+\alpha_\delta)-1} \cdot \exp(-(\lambda^2 + \beta_\delta)\delta). \end{aligned}$$

Then,

$$\delta \mid \boldsymbol{Y}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2, \lambda^2, r \sim Gamma(r + \alpha_\delta, \lambda^2 + \beta_\delta).$$

A.2 Bayesian Lasso in Circular-(Circular, Linear) Regression Models

A.2.1 Classical Bayesian Lasso

Assume $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_n]^T$ is a $n \times 1$ vector of circular response. Let $\Omega_2 = \{\mu_\theta, \boldsymbol{\beta}, \kappa, \boldsymbol{\tau}^2, \lambda^2\}$ be parameter space, where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_m)^T$ is $m \times 1$ vector of regression coefficients and $\boldsymbol{\tau}^2 = (\tau_1^2, \tau_2^2 ..., \tau_m^2)$, with m denoting the total number of circular and linear covariates. Then the joint pdf is given by:

$$p(\boldsymbol{\theta}, \Omega_2) = L(\mu_\theta, \boldsymbol{\beta}, \kappa \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2) \cdot p(\kappa \mid a, b) \cdot p(\boldsymbol{\tau}^2 \mid \lambda^2) \cdot p(\lambda^2 \mid c, d) \cdot p(\mu_\theta)$$

Here's the detailed breakdown of each component:

- **Likelihood**:

$$\begin{aligned} L(\mu_\theta, \boldsymbol{\beta}, \kappa \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{e^{\kappa \cos(\theta_i - \mu_\theta - g(\boldsymbol{\phi}_{hi}\boldsymbol{\beta}_{\boldsymbol{\phi}_h} + \boldsymbol{X}_{ki}\boldsymbol{\beta}_{\boldsymbol{X}_k}))}}{2\pi I_0(\kappa)}, \\ 0 \leq \theta_i < 2\pi, 0 \leq \mu_\theta < 2\pi, \kappa > 0, \\ h = 1, 2, ..., p,\ k = 1, 2, ..., q,\ i = 1, 2, ..., n, \end{aligned}$$

where μ_θ is mean direction and κ is concentration parameter.

- **Prior for $\boldsymbol{\beta}$**:

$$p(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2) = \frac{1}{(2\pi)^{m/2} \mid \mathbf{D}_\tau \mid^{1/2}} e^{-\frac{1}{2}\boldsymbol{\beta}^T \mathbf{D}_\tau{}^{-1}\boldsymbol{\beta}}, -\infty \leq \boldsymbol{\beta} < \infty,$$

where $\mathbf{D}_\tau = diag(\tau_1^2, ..., \tau_m^2)$.

- **Prior for κ**:

$$p(\kappa) = \frac{b^a}{\Gamma(a)} \kappa^{a-1} e^{-b\kappa}, \kappa > 0, a > 0, b > 0,$$

where a and b are shape and rate parameters, respectively.

- **Priors for τ_j^2**: For each $j = 1, \ldots, m$:

$$p(\tau_j^2) = \frac{\lambda^2}{2} e^{-\frac{\lambda^2}{2}\tau_j^2}, \tau_j^2 > 0, \lambda^2 > 0,$$

where λ^2 is rate parameter.

- **Prior for λ^2**:

$$p(\lambda^2) = \frac{d^c}{\Gamma(c)} (\lambda^2)^{c-1} e^{-d\lambda^2}, \lambda^2 > 0, c > 0, d > 0,$$

where c and d are shape and rate parameters, respectively.

- **Prior for μ_θ**:

$$p(\mu_\theta) = \frac{1}{2\pi}, \quad \text{for } 0 \leq \mu_\theta < 2\pi.$$

We derive the full conditional distribution for each parameter.

- **Full conditional distribution for μ_θ**:

  The full conditional distribution of μ_θ given all other parameters is proportional to the product of the likelihood terms involving μ_θ:

$$\begin{aligned} p(\mu_\theta \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \kappa, \boldsymbol{\tau}^2, \lambda^2) &\propto L(\mu_\theta, \boldsymbol{\beta}, \kappa \mid \boldsymbol{\theta}) \\ &\propto \prod_{i=1}^{n} e^{\kappa \cos(\theta_i - \mu_\theta - g(\boldsymbol{\phi}_{hi}\boldsymbol{\beta}_{\boldsymbol{\phi}_h} + \boldsymbol{X}_{ki}\boldsymbol{\beta}_{\boldsymbol{X}_k}))}. \end{aligned}$$

- **Full conditional distribution for $\boldsymbol{\beta}$**:

  The full conditional distribution for $\boldsymbol{\beta}$ given all other parameters can be derived by combining the likelihood and prior:

$$\begin{aligned} p(\boldsymbol{\beta} \mid \boldsymbol{\theta}, \mu_\theta, \kappa, \boldsymbol{\tau}^2, \lambda^2) &\propto L(\mu_\theta, \boldsymbol{\beta}, \kappa \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2) \\ &\propto \prod_{i=1}^{n} e^{\kappa \cos(\theta_i - \mu_\theta - g(\boldsymbol{\phi}_{hi}\boldsymbol{\beta}_{\boldsymbol{\phi}_h} + \boldsymbol{X}_{ki}\boldsymbol{\beta}_{\boldsymbol{X}_k}))} \cdot \exp(-\frac{1}{2}\boldsymbol{\beta}^T D\tau^{-1}\boldsymbol{\beta}). \end{aligned}$$

- **Full conditional distribution for κ**:

  The full conditional distribution for κ given all other parameters is given by:

$$\begin{aligned} p(\kappa \mid \boldsymbol{\theta}, \mu_\theta, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \lambda^2\}) &\propto L(\mu_\theta, \boldsymbol{\beta}, \kappa \mid \boldsymbol{\theta}) \cdot p(\kappa \mid a, b) \\ &\propto \prod_{i=1}^{n} \frac{e^{\kappa \cos(\theta_i - \mu_\theta - g(\boldsymbol{\phi}_{hi}\boldsymbol{\beta}_{\boldsymbol{\phi}_h} + \boldsymbol{X}_{ki}\boldsymbol{\beta}_{\boldsymbol{X}_k}))}}{2\pi I_0(\kappa)} \cdot \kappa^{a-1} e^{-b\kappa}. \end{aligned}$$

- **Full conditional distribution for τ_j^2, $j = 1, \ldots, m$**:

  The full conditional distribution for τ_j^2 given all other parameters is given by:

$$\begin{aligned} p(\tau_j^2 \mid \boldsymbol{\theta}, \mu_\theta, \beta_j, \kappa, \tau_{-j}^2, \lambda^2) &\propto p(\beta_j \mid \tau_j^2) \cdot p(\tau_j^2 \mid \lambda^2) \\ &\propto (\tau_j^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_j^2} + \lambda^2 \tau_j^2)} \\ &\propto (1/\tau_j^2)^{-\frac{3}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_j^2} + \frac{\lambda^2}{1/\tau_j^2})} \\ &\propto (1/\tau_j^2)^{-\frac{3}{2}} \cdot e^{-\frac{\beta_j^2(\frac{1}{\tau_j^2} - \sqrt{\lambda^2/\beta_j^2})^2}{2(1/\tau_j^2)}}. \end{aligned}$$

  This is in the form of the reciprocal of the Inverse-Gaussian distribution. Then

$$\frac{1}{\tau_j^2} \mid \boldsymbol{\theta}, \mu_\theta, \beta_j, \kappa, \tau_{-j}^2, \lambda^2 \sim Inverse - Gaussian(\sqrt{\frac{\lambda^2}{\beta_j^2}}, \lambda^2).$$

- **Full conditional distribution for λ^2**:

  The full conditional distribution for λ^2 given all other parameters is given by:

$$\begin{aligned} p(\lambda^2 \mid \boldsymbol{\theta}, \mu_\theta, \boldsymbol{\beta}, \kappa, \boldsymbol{\tau}^2) &\propto p(\tau_1^2, ..., \tau_m^2 \mid \lambda^2) \cdot p(\lambda^2 \mid c, d) \\ &\propto \prod_{j=1}^{m} \frac{\lambda^2}{2} \cdot e^{-\frac{\lambda^2 \tau_j^2}{2}} (\lambda^2)^{c-1} e^{-d\lambda^2} \\ &\propto (\lambda^2)^{m+c-1} \cdot e^{-\lambda^2(\frac{1}{2}\sum_{j=1}^{m} \tau_j^2 + d)}. \end{aligned}$$

  We can recognize this as a gamma density for λ^2:

$$\lambda^2 \mid \boldsymbol{\theta}, \mu_\theta, \boldsymbol{\beta}, \kappa, \boldsymbol{\tau}^2 \sim Gamma(m + c, \frac{1}{2}\sum_{j=1}^{m} \tau_j^2 + d).$$

A.2.2 Bayesian Adaptive Lasso

The likelihood function and prior distributions defined in this method are the same as those defined in the classical Bayesian lasso method for all parameters except for τ_j^2 and λ^2. Therefore, the full conditional distributions for all parameters except τ_j^2 and λ^2 are the same as those in the classical Bayesian lasso method. This method has λ_j^2 instead of λ^2. Accordingly, prior distributions for τ_j^2 and λ_j^2 are as follows:

- **Priors for** $\tau_j^2, j = 1, \ldots, m$:

$$p(\tau_j^2 \mid \lambda_j^2) = \frac{\lambda_j^2}{2} e^{-\frac{\lambda_j^2}{2}\tau_j^2},\ \tau_j^2 > 0,\ \lambda_j^2 > 0,$$

  where λ_j^2 is rate parameter.

- **Priors for** $\lambda_j^2, j = 1, \ldots, m$:

$$p(\lambda_j^2 \mid c, d) = \frac{d^c}{\Gamma(c)} (\lambda_j^2)^{c-1} e^{-d\lambda_j^2},\ \lambda_j^2 > 0,\ c > 0,\ d > 0,$$

  where c and d are shape and rate parameters, respectively.

Based on these prior distributions, we derive the full conditional distributions for τ_j^2 and λ_j^2.

- **Full conditional distribution for** $\tau_j^2, j = 1, \ldots, m$:

  The full conditional distribution for τ_j^2 given all other parameters is given by:

$$\begin{aligned} p(\tau_j^2 \mid \boldsymbol{\theta}, \mu_\theta, \beta_j, \kappa, \tau_{-j}^2, \lambda_j^2) &\propto p(\beta_j \mid \tau_j^2) \cdot p(\tau_j^2 \mid \lambda_j^2) \\ &\propto (\tau_j^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_j^2} + \lambda_j^2 \tau_j^2)} \\ &\propto (1/\tau_j^2)^{-\frac{3}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_j^2} + \frac{\lambda_j^2}{1/\tau_j^2})} \\ &\propto (1/\tau_j^2)^{-\frac{3}{2}} \cdot e^{-\frac{\beta_j^2(\frac{1}{\tau_j^2} - \sqrt{\lambda_j^2/\beta_j^2})^2}{2(1/\tau_j^2)}}. \end{aligned}$$

  This is in the form of the reciprocal of the Inverse-Gaussian distribution. Then

$$\frac{1}{\tau_j^2} \mid \boldsymbol{\theta}, \mu_\theta, \beta_j, \kappa, \tau_{-j}^2, \lambda_j^2 \sim Inverse - Gaussian(\sqrt{\frac{\lambda_j^2}{\beta_j^2}}, \lambda_j^2).$$

- **Full conditional distribution for λ_j^2, $j = 1, \ldots, m$:**

  The full conditional distribution for λ_j^2 given all other parameters is given by:

$$\begin{aligned} p(\lambda_j^2 \mid \boldsymbol{\theta}, \mu_\theta, \boldsymbol{\beta}, \kappa, \boldsymbol{\tau^2}, \lambda_{-j}^2) &\propto p(\tau_1^2, ..., \tau_m^2 \mid \lambda_j^2) \cdot p(\lambda_j^2 \mid c, d) \\ &\propto \frac{\lambda_j^2}{2} \cdot e^{-\frac{\lambda_j^2 \tau_j^2}{2}} (\lambda_j^2)^{c-1} e^{-d\lambda_j^2} \\ &\propto (\lambda_j^2)^{(c+1)-1} \cdot e^{-\lambda_j^2(\frac{1}{2}(\tau_j^2+d))}. \end{aligned}$$

  We can recognize this as a gamma density for λ_j^2:

$$\lambda_j^2 \mid \boldsymbol{\theta}, \mu_\theta, \boldsymbol{\beta}, \kappa, \boldsymbol{\tau^2}, \lambda_{-j}^2 \sim Gamma(c+1, \frac{1}{2}\tau_j^2 + d).$$

A.2.3 Dirichlet Lasso 1

In this method, unlike previous methods, Dirichlet process prior with the mass parameter α and baseline measure G_0 is defined for tuning parameters ($\lambda_j' s$). If baseline measure G_0 is set to a usual prior distribution such as gamma, the full conditional distributions align with those in the preceding methods. Setting G_0 to gamma distribution establishes a consistent framework for further analysis and comparison.

The DP prior can be constructed using the stick-breaking construction, which is a common way to intuitively understand and generate samples from the DP distribution.

Mathematically, the stick-breaking process can be described as follows:

- Choose stick-breaking weights: $v_1, v_2, v_3, \ldots$ where $v_i \sim$ Beta$(1, \alpha)$, where α is a concentration parameter.
- The cluster probabilities $p_1, p_2, p_3, \ldots$ can be derived from the stick-breaking weights: letting $p_1 = v_1$ and $p_i = v_i \prod_{j=1}^{i-1}(1 - v_j)$ for $i = 2$ to ∞.
- The resulting mixture distribution can be written as:

$$G \sim \sum_{i=1}^{\infty} p_i \delta(\psi_i),$$

  where $\delta(\psi_i)$ is a point mass distribution at ψ_i, where $\psi_i \sim G_0$. The possibility of achieving a precise estimation of G by limiting the infinite series to a finite

integer C is demonstrated in the study of Ishwaran and James (2002). Comprehensive information regarding the method of estimation and complete conditional probability distributions for the model parameters are available therein. We use the stick-breaking representation of the DP in this thesis. Accordingly, the full conditional distributions are as follows:

- **The conditional distribution for $\boldsymbol{p}$**:

$$p_1 = v_1^* \text{ and } p_k = (1 - v_1^*)(1 - v_2^*)\dots(1 - v_{k-1}^*)v_k^*, \quad k = 2, \dots, C-1$$
$$v_k^* \sim \text{Beta}\left(1 + n_k, \alpha + \sum_{l=k+1}^{C} n_l\right), \quad k = 1, \dots, C-1.$$

  Here, n_k^* is the number of data points assigned to component k.

- **The full conditional distribution for α**:

  If a gamma prior is assigned for α, i.e. $\alpha \sim Gamma(\eta_1, \eta_2)$, the full conditional distribution is as follows;

$$\alpha \sim \text{Gamma}\left(C + \eta_1 - 1, \eta_2 - \sum_{k=1}^{C-1} \log(1 - v_k^*)\right)$$

  where v_k^* represents the same values in the simulation of $\boldsymbol{p}$.

The remaining full conditional distributions follow a similar pattern as those observed in the BaLasso.

A.2.4 Dirichlet Lasso 2

The likelihood function and prior distributions established in this method coincide with those defined in the Dirichlet Lasso 1 method for all parameters except for $\boldsymbol{\beta}$. Hence, the full conditional distributions for all parameters except $\boldsymbol{\beta}$ match those present in the Dirichlet Lasso 1. Here, prior distributions for $\boldsymbol{\beta}$ and additional parameter such as $\boldsymbol{W}$ and $\boldsymbol{\pi}$ are as follows:

- **Prior for $\boldsymbol{\beta}$**:

$$p(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \boldsymbol{W}) = (1 - \boldsymbol{W}) f(*) + (\boldsymbol{W}) \frac{1}{(2\pi)^{m/2} \mid \mathbf{D}_\tau \mid^{1/2}} e^{-\frac{1}{2}\boldsymbol{\beta}^T \mathbf{D}_\tau{}^{-1} \boldsymbol{\beta}}$$

  where $f(*)$ denotes the point mass at 0, $\boldsymbol{W} = (W_1, W_2 \ldots W_m)$ denotes a vector of binary random variable with $P(W_j = 1) = \pi_j$.

- **Priors for** W_j, $j = 1, \ldots, m$:

$$p(W_j \mid \pi_j) = \pi_j^{w_j} \cdot (1 - \pi_j)^{1 - w_j}, \; w_j = 0 \quad \text{or} \quad 1 \quad \text{for } 0 \leq \pi_j < 1.$$

- **Priors for** π_j, $j = 1, \ldots, m$:

$$p(\pi_j \mid \gamma_1, \gamma_2) = \frac{1}{B(\gamma_1, \gamma_2)} \cdot \pi_j^{\gamma_1 - 1} \cdot (1 - \pi_j)^{\gamma_2 - 1},$$

  where $0 \leq \pi_j \leq 1$, γ_1 and γ_2 are the shape parameters of the distribution where $\gamma_1 > 0$ and $\gamma_2 > 0$.

The full conditional distributions are as follows:

- **Full conditional distribution for $\boldsymbol{\beta}$**:

$$p(\boldsymbol{\beta} \mid \boldsymbol{\theta}, \mu_\theta, \kappa, \boldsymbol{\tau}^2, \boldsymbol{W}, \boldsymbol{\lambda^2}) \propto L(\mu_\theta, \boldsymbol{\beta}, \kappa \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\beta} \mid \boldsymbol{\tau}^2, \boldsymbol{W})$$

  - If $W_j = 0$ then $\beta_j = 0$ (point mass at 0).
  - If $W_j = 1$ then the full conditional distribution is the same as in other methods.

- **Full conditional distribution for $\boldsymbol{W}$**: The expression for the full conditional distribution of W_j, $j = 1, \ldots, m$ is given by:

$$\begin{aligned}
&p(W_j \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \mu_\theta, \kappa, \boldsymbol{\tau^2}, W_{-j}, \boldsymbol{\pi}) \propto p(\beta_j \mid \tau_j^2, W_j) \cdot p(W_j \mid \pi_j) \\
&\propto [(w_j \cdot (\tau_j^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_j^2})}) + ((1 - w_j) \cdot (\tau_j^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_j^2})}] \cdot \pi_j^{w_j} (1 - \pi_j)^{w_j}) \\
&\propto [(\tau_j^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_j^2})}] \cdot \pi_j^{w_j} (1 - \pi_j)^{w_j}).
\end{aligned}$$

- **Full conditional distribution for $\boldsymbol{\pi}$**: The full conditional distribution of π_j, $j = 1, \ldots, m$ is given by:

$$\begin{aligned} p(\pi_j \mid \boldsymbol{\theta}, \mu_\theta, \boldsymbol{\beta}, \kappa, \boldsymbol{\tau^2}, \boldsymbol{\lambda}^2, W_j, \pi^2_{-j}) &\propto p(W_j \mid \pi_j) \cdot p(\pi_j \mid \gamma_1, \gamma_2) \\ &\propto \pi_j^{w_j} \cdot (1-\pi_j)^{1-w_j} \cdot \pi_j^{\gamma_1 - 1} \cdot (1-\pi_j)^{\gamma_2 - 1} \\ &\propto \pi_j^{w_j + \gamma_1 - 1} \cdot (1-\pi_j)^{1-w_j+\gamma_2-1} \end{aligned}$$

This is proportional to the kernel of a Beta distribution with updated parameters $w_j + \gamma_1$ and $1 - w_j + \gamma_2$. So, the full conditional distribution for π_j is a Beta distribution as follows:

$$\pi_j \mid \boldsymbol{\theta}, \mu_\theta, \boldsymbol{\beta}, \kappa, \boldsymbol{\tau^2}, \boldsymbol{\lambda}^2, \boldsymbol{W}, \pi^2_{-j} \sim \text{Beta}(w_j + \gamma_1, 1 - w_j + \gamma_2).$$

A.2.5 Dirichlet Mixture Lasso

Unlike the Dirichlet lasso 1, $\boldsymbol{\tau}_2$ and $\boldsymbol{\lambda}_2$ parameters are added in this method and the prior distribution of $\boldsymbol{\beta}$ is a mixture of two normal distributions. Accordingly, the full conditional distributions are derived and as follows:

- **Full conditional distribution of $\boldsymbol{\beta}$:**

The full conditional distribution for $\boldsymbol{\beta}$ given all other parameters:

$$\begin{aligned} &p(\boldsymbol{\beta} \mid \boldsymbol{\theta}, \mu_\theta, \kappa, \boldsymbol{\tau}_1^2, \boldsymbol{\tau}_2^2, \boldsymbol{\pi}) \propto L(\mu_\theta, \boldsymbol{\beta}, \kappa \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\beta} \mid \boldsymbol{\tau}_1^2, \boldsymbol{\tau}_2^2, \boldsymbol{\pi}) \\ &\propto \prod_{i=1}^{n} e^{\kappa \cos(\theta_i - \mu_\theta - g(\boldsymbol{\phi}_{hi}\boldsymbol{\beta}_{\boldsymbol{\phi}_h} + \boldsymbol{X}_{ki}\boldsymbol{\beta}_{\boldsymbol{X}_k}))} \cdot [\boldsymbol{\pi} \cdot \exp(-\frac{1}{2}\boldsymbol{\beta}^T D \tau_1^{-1} \boldsymbol{\beta}) + \\ &(\mathbf{1} - \boldsymbol{\pi}) \cdot \exp(-\frac{1}{2}\boldsymbol{\beta}^T D \tau_2^{-1} \boldsymbol{\beta})]. \end{aligned}$$

- **Full conditional distribution of τ_{1j}^2 and τ_{2j}^2 :**

For each $j = 1, \ldots, m$, the full conditional distribution for τ_{1j}^2 given all other parameters is given by:

$$p(\tau_{1j}^2 \mid \boldsymbol{\theta}, \mu_\theta, \kappa, \beta_j, \tau_{-1j}^2, \tau_{2j}^2, \boldsymbol{\pi}) \propto p(\beta_j \mid \tau_{1j}^2, \tau_{2j}^2, \pi_j) \cdot p(\tau_{1j}^2 \mid \lambda_{1j}^2)$$
$$\propto [(\pi_j \cdot (\tau_{1j}^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_{1j}^2})}) + ((1-\pi_j) \cdot (\tau_{2j}^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_{2j}^2})}] \cdot e^{-\frac{1}{2}\lambda_{1j}^2 \tau_{1j}^2}).$$

Similarly for τ_{2j}^2 is given by:

$$p(\tau_{2j}^2 \mid \boldsymbol{\theta}, \mu_\theta, \kappa, \beta_j, \tau_{1j}^2, \tau_{-2j}^2, \boldsymbol{\pi}) \propto p(\beta_j \mid, \tau_{1j}^2, \tau_{2j}^2, \pi_j) \cdot p(\tau_{2j}^2 \mid \lambda_{2j}^2)$$
$$\propto [(\pi_j \cdot (\tau_{1j}^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_{1j}^2})}) + ((1-\pi_j) \cdot (\tau_{2j}^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_{2j}^2})}] \cdot e^{-\frac{1}{2}\lambda_{2j}^2 \tau_{2j}^2}.$$

- **Full conditional distribution of λ_{1j}^2 and λ_{2j}^2:**

  If baseline measure G_{01} and G_{01} are set to a usual prior distribution such as gamma, the full conditional distributions will align with those in the preceding methods.

- **Full conditional distribution for $\boldsymbol{\pi}$**:

  The full conditional distribution of π_j is given by:

$$p(\pi_j \mid \boldsymbol{\theta}, \beta_j, \mu_\theta, \kappa, \tau_{1j}^2, \tau_{2j}^2, \pi_{-j}^2) \propto p(\beta_j \mid, \tau_{1j}^2, \tau_{2j}^2, \pi_j) \cdot p(\pi_j \mid \gamma_3, \gamma_4)$$
$$\propto [(\pi_j \cdot (\tau_{1j}^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_{1j}^2})}) + ((1-\pi_j) \cdot (\tau_{2j}^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\frac{\beta_j^2}{\tau_{2j}^2})}] \cdot \pi_j^{\gamma_3}(1-\pi_j)^{\gamma_4}).$$

The rest of the full conditional distributions are the sames as those in BaLasso and $DLasso_1$.

Note that some of these distributions do not have simple closed-form expressions, and we need to use techniques like MCMC methods to sample from these distributions.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Çamlı, Onur

EDUCATION

Degree	Institution	Year of Graduation
MS	METU, Department of Statistics	2017
BS	ESOGU, Department of Statistics	2012

PROFESSIONAL EXPERIENCE

Year	Institution	Position
2013-2014	ETU, Department of Statistics	Research Assistant
2014-2023	METU, Department of Statistics	Research Assistant

PUBLICATIONS

Camli, O., Kalaylioglu, Z., and SenGupta, A. (2022). Variable selection in linear-circular regression models. Journal of Applied Statistics, 1-25.

Sareyyüpoğlu, B., Gülyaz, G., Saraç, F., Uzar, S., Kabaklı, Ö., Çokçalışkan, C., ... and Camli, O. (2022). Antibody response of sheep to simultaneous vaccination of foot-and-mouth disease, peste des petits ruminants, sheep pox and goat pox, and blue-tongue vaccines. Tropical Biomedicine, 39(1), 47-54.

Camli, O., and Kalaylioglu, Z. (2021). Bayesian predictive model selection in circular random effects models with applications in ecological and environmental studies. Environmental and Ecological Statistics, 28, 21-34.