Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Variable selection in linear-circular regression models

Onur Camli, Zeynep Kalaylioglu & Ashis SenGupta

JOURNAL OF APPLIED STATISTICS
https://doi.org/10.1080/02664763.2022.2110860

Variable selection in linear-circular regression models

Onur Camli [a], Zeynep Kalaylioglu [a] and Ashis SenGupta [b]

[a]Department of Statistics, Middle East Technical University, Ankara, Türkiye; [b]Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

ABSTRACT

Applications of circular regression models are ubiquitous in many disciplines, particularly in meteorology, biology and geology. In circular regression models, variable selection problem continues to be a remarkable open question. In this paper, we address variable selection in linear-circular regression models where uni-variate linear dependent and a mixed set of circular and linear independent variables constitute the data set. We consider Bayesian lasso which is a popular choice for variable selection in classical linear regression models. We show that Bayesian lasso in linear-circular regression models 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 while using Gibbs Sampling. This hyper-prior construction is computationally more feasible than the hyper-priors that are based on correlation measures. We show in a comprehensive simulation study that Bayesian lasso with EB-GS hyper-prior leads to a more robust inference. Overall, the method offers an efficient Bayesian lasso for variable selection in linear-circular regression while reducing model complexity.

ARTICLE HISTORY
Received 11 November 2021
Accepted 1 August 2022

KEYWORDS
Regularization; Bayesian lasso; laplace distribution; circular regression; dimension reduction

1. Introduction

Circular variable is a variable that can be mapped onto a unit circle such as time of the day and month of the year as well as angular or cardinal directions. Circular regression models are used in many different fields such as meteorology, biology, geology, medicine, and psychology when at least one of the variables of interest is circular. Circular regression examples include air quality index on wind speed and direction, directional behavior of sandhoppers depending on some environmental factors, the effect of tidal characteristics of the fish's environment on the spawning time [3,16,20]. The theoretical and methodological aspects of circular regression models are addressed in [6,10,16,18,30]. The focus of this paper is linear-circular regression models. Linear-circular regression models are used when the response is a linear variable whereas at least one of the covariates is circular [16].

CONTACT Onur Camli camli@metu.edu.tr

As far as the regression models are concerned, one of the prevailing problems is variable selection, both in linear and circular context. In the circular context, there have been an array of variable/model selection methods from both frequentist and Bayesian perspectives [3,4,6,20,27,29,32] but it is still an active research area. We believe lasso [35] that has been proven successful in the linear context would be beneficial in the circular context. In this paper, we consider lasso in variable selection in linear-circular regression models.

There are many advantages of lasso in linear regression models. Primary advantages are good prediction accuracy at the cost of negligible bias, satisfactorily parsimonious models, controlled risk of overfitting, increased model interpretability, computational benefits, straightforward adaptation in wide range of models [34], and readily available Bayesian interpretation [35]. In this paper, we provide an adaptation of Bayesian lasso for variable selection in linear-circular regression models motivated by the performance of lasso in linear case. We developed a new prior construction for the tuning parameter based on unification of notions of Empirical Bayes (EB) and Gibbs Sampling (GS) which we call EB-GS prior. Rest of the paper is organized as follows. Section 2 presents an overview of the lasso and Bayesian lasso in linear regression models. We devote Section 3 to tuning parameter specification and its role in lasso. In this section we also introduce a new alternative EB type approach for specifying hyper-hyper parameters of hyper-prior distribution on the tuning parameter. Section 4 introduces the adaptation of Bayesian lasso in linear-circular regression models and discusses some specific properties such as posterior consistency and asymptotic normality. Section 5 presents extensive simulation studies for assessing the sensitivity and the performance of the proposed method. In Section 6, the method is applied to two datasets that are important in wildfire studies. Finally, Section 7 completes the paper with some concluding remarks and some future works related to the proposed method.

2. Review of the lasso

Consider the following linear regression model:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \tag{1}$$

where $\mathbf{Y}_{n\times 1}$ is a vector of linear response, $\mathbf{X}_{n\times(p+1)}$ is a matrix of linear covariates, $\beta_{(p+1)\times 1}$ is a vector of unknown regression coefficients and $\epsilon_{n\times 1}$ is a vector of random error term. Here, n and p denote the sample size and the number of linear covariates, respectively. The most popular estimation method has been ordinary least squares (OLS) for model in Equation (1). When we use the OLS estimates, two challenges may arise, (i) prediction accuracy, (ii) interpretability [35]. If the purpose of modeling is prediction, we naturally would like our model to accurately predict future data. Although the OLS estimates often have low bias, they have large variance somehow leading to a loss in prediction accuracy. Prediction accuracy can often be improved by shrinking the regression coefficients. In this framework, 'shrinkage' refers to decreasing the magnitude of regression coefficients towards zero. Shrinkage sacrifices some unbiasedness to reduce the variance of the predicted value and may improve the overall prediction as a result of this compromise. OLS cannot distinguish variables with little or no influence. Practitioners would like to pick only a subset of all variables that are assumed to be relevant. If shrinkage is sufficiently large, some of the regression coefficients are driven to zero, leading to a sparse model which can be interpreted more easily.

Tibshirani [35] proposed a technique accomplishing this, called the lasso (least absolute shrinkage and selection operator) for model in Equation (1). The lasso coefficients are solutions to the l_1 optimization problem and the lasso estimate $\hat{\beta}^{\text{lasso}}$ is defined by

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

or equivalently [24]

$$\hat{\beta}^{\text{lasso}} = \arg\min\{||Y - X\beta||_2^2 + \lambda||\beta||_1\}, \tag{3}$$

where t $\geq$ 0 and $\lambda \geq 0$ are tuning parameters, $||\beta||_1$ is the lasso penalty term. 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 sufficiently large values of 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). To arrive at a final lasso estimator of β, the lasso penalty parameter λ needs to be chosen.

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 [35] described some efficient and stable algorithms for the solution of this problem and so the lasso of Tibshirani has become a widely used alternative to OLS in regression problems.

The lasso has an alternative Bayesian interpretation. Tibshirani [35] stated that the lasso estimates can be obtained by the mode of the conditional posterior distribution of the regression coefficients in a Bayesian model in which each coefficient is assigned apriori the same double exponential (DE) or Laplace densities. When a DE or Laplace prior with location $\mu = 0$ and scale $\frac{1}{\lambda}$ is assumed for β, i.e.

$$\pi(\beta_j|\lambda) = \frac{\lambda}{2}\, e^{-\lambda|\beta_j|}, \quad j = 1, \ldots, p \tag{4}$$

where π denotes the prior density, $\lambda > 0$, and p is the number of covariates, then the joint posterior distribution of regression coefficients, β, is proportional to the following quantity

$$f(\beta|\mathbf{X}, \sigma^2, \lambda) \propto \exp\left(-\frac{1}{\sigma^2}||\mathbf{Y} - \mathbf{X}\beta||_2^2 + \lambda||\beta||_1\right). \tag{5}$$

The equivalency between two objectives in Equation (3) and in Equation (5) can be seen easily.

In Bayesian interpretation of the lasso, the objective function is interpreted as the negative log-likelihood function, the lasso penalty term is seen as the negative log-prior distribution of regression coefficients and the lasso estimates are the global maxima of the posterior distribution. In other words, Bayesian interpretation of the lasso is considered as maximum aposteriori under double-exponential prior.

3. 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 [13,35,38], (ii) information criterion-based approaches such as AIC, BIC, and generalized information criterion [2,8,9,11,15,34,36,37], (iii) model metrics-based methods such as the residual sum of squares or Mallows's Cp-type selection criterion[7,15], (iv) an analytical unbiased estimate of risk (Stein's unbiased risk estimation) [35].

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 [25], (ii) Bayes factor [21], (iii) selecting an appropriate hyper-prior [14,19,21,22,25,28], (iv) information criterion-based approaches such as DIC [17]. The focus of the current article 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 [25] considered the class of gamma hyper-priors on λ^2 which is given by:

$$\pi(\lambda^2) = \frac{\delta^r}{\Gamma(r)}(\lambda^2)^{r-1}\,\mathrm{e}^{-\delta\lambda^2},\ \lambda^2 > 0,\ r > 0,\ \delta > 0, \tag{6}$$

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 5.1, 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 [25] 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 Ntzoufras [21]. Their approach gives similar inclusion probabilities for significant covariates to those of a non-informative prior (such as $Ga(0.01, 0.01)$).

Table 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)

3.1. 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 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 in Table 1 and explained 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.

4. Bayesian lasso in linear-circular regression models

In this study, we focus on linear-circular regression models and their cosine representation as a popular choice of modeling [18]. Cosine model is linear in both the variables and the unknown regression coefficients. Here we adapt our EB-GS method for this model for robust variable selection.

4.1. Cosine model

Suppose one observes $(y_i, \theta_i)_{i=1}^{n}$, where $y_i \in (-\infty, \infty)$ and $\theta_i \in [0, 2\pi]$ are linear and circular measurements for i=1,..., n. For the regression of a linear random variable Y on a circular variable θ, we use the model given by [18], that is

$$Y_i = \alpha + \beta \cos(\theta_i - \mu_\theta) + \epsilon_i, \quad i = 1, \ldots, n \tag{7}$$

where, α and β are the intercept and slope parameters respectively, μ_θ 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 (7) as follows.

$$Y_i = \beta_0 + \beta_1 \cos\theta_i + \beta_2 \sin\theta_i + \epsilon_i, \quad i = 1, \dots, n. \tag{8}$$

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 each 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 (8) can be estimated by minimizing the following sum of squared distances function $Q(.)$ given by

$$Q(\Omega) = \arg\min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1\cos\theta_i - \beta_2\sin\theta_i)^2 \text{ subject to } \sum_{j=1}^{2} |\beta_j| \le t. \tag{9}$$

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

The model in Equation (8) can be extended as follows for p independent circular covariates,

$$Y_i = \beta_0 + \sum_{k=1}^{p} \{\beta_{1k}\cos\theta_{ik} + \beta_{2k}\sin\theta_{ik}\} + \epsilon_i, \quad i = 1, \dots, n, \quad k = 1, \dots, p \tag{10}$$

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 (10) can be estimated by minimizing the following sum of squared distances,

$$Q(\Omega) = \arg\min \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{k=1}^{p} \{\beta_{1k}\cos\theta_{ik} + \beta_{2k}\sin\theta_{ik}\})^2 \text{ subject to } \sum_{k=1}^{p}\sum_{j=1}^{2} |\beta_{jk}| \le t. \tag{11}$$

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

4.2. Adaptation of Bayesian lasso

Combining the model structure we give in Equation (10) with the alternative representation for DE of lasso given by [1,25], we have the following hierarchical structure

$$\begin{aligned}
\mathbf{Y}|\beta_0, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim N(\beta_0 + \mathbf{X}\boldsymbol{\beta}, \sigma^2 I),\\
\boldsymbol{\beta}|\tau_1^2, \dots, \tau_{2p}^2, \sigma^2 &\sim N(0, \sigma^2\mathbf{D}_\tau),\ \mathbf{D}_\tau = diag(\tau_1^2, \dots, \tau_{2p}^2),\\
\sigma^2 &\sim Inv - Ga(a, b),\\
\tau_j^2 &\sim Exp(\lambda^2/2), \quad j = 1, \dots, 2p,\\
\lambda^2 &\sim Ga(r, \delta),\\
\beta_0 &\sim f(\beta_0).
\end{aligned} \tag{12}$$

where $\mathbf{X} = (X_1, X_2, \dots, X_{2p-1}, X_{2p})$ is the vector of covariates with $X_1 = \cos(\theta_1)$, $X_2 = \sin(\theta_1), \dots, X_{2p-1} = \cos(\theta_p), X_{2p} = \sin(\theta_p)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_{2p-1}, \beta_{2p})$ is the vector of regression coefficients, with $\beta_1 = \beta_{11}, \beta_2 = \beta_{21},\ldots, \beta_{2p-1} = \beta_{1p}, \beta_{2p} = \beta_{2p}, f(\beta_0)$ 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,

$$
\begin{aligned}
&\beta_0 \sim N(\bar{y}, \sigma^2/n),\\
&\boldsymbol{\beta} \sim N(\mathbf{A^{-1}X^{T}y}, \sigma^2\mathbf{A^{-1}}),\ \mathbf{A} = \mathbf{X^{T}X} + \mathbf{D}_{\tau}^{-1},\\
&\sigma^2 \sim Inv-Ga((n-1)/2 + p/2, ((\boldsymbol{y} - \beta_0 - \boldsymbol{X\beta})^{T}(\boldsymbol{y} - \beta_0 - \boldsymbol{X\beta}) + \boldsymbol{\beta}^{T}\boldsymbol{D}_{\tau}^{-1}\boldsymbol{\beta})/2 + b),\\
&\frac{1}{\tau_j^2} \sim Inv-Gauss(\sqrt{\frac{\lambda^2\sigma^2}{\beta_j^2}}, \lambda^2),\\
&\lambda^2 \sim Ga(p + r, \sum_{j=1}^{2p}\tau_j^2/2 + \delta).
\end{aligned}
\tag{13}
$$

We used posterior mean as Bayesian lasso estimates. MCMC performed without a 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.

5. Simulation study

This section first investigates the sensitivity of the posterior results to standard λ hyperpriors and EB-GS hyper-prior (Section 5.1). Then, it investigates the performance of the proposed EB-GS method in parameter estimation and variable selection (Section 5.2). All this is accomplished by orchestrating a comprehensive Monte Carlo (MC) study.

Our simulation scenarios mimic those in the original lasso paper [35]. 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 (vM) distributions with circular mean 0 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 4.1 with $\epsilon_i \sim N(0, 9)$. The study is controlled for the sample size using, $n = 100, 250$.

The three simulation scenarios considered are as follows:

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

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 $\sigma^2 \sim Ga(0.001, 0.001)$. The data are generated using *rvonmises* function in R language. In order to implement MCMC scheme, OpenBUGS which is an open source software for Bayesian statistics is employed. R programming language and OpenBUGS program are integrated to carry out all analyses in this study. Trace plots and Brooks-Gelman-Rubin (BGR) statistics [33] are used to monitor convergence as well as to determine warm-up period and number of MCMC samples to be used for final posterior inference. Each scenario is repeated 500 times.

Performance in variable selection is measured by the average number of covariates that are correctly included in the model. Of the covariates that are found significant, the ones that were used in the true response data generating mechanism are considered correctly included in the model. A covariate is deemed significant if either one of the Bayesian credible interval (CI) for coefficient of cosine or sine term excludes nullity. Specific definitions and calculations of these measures are as follows;

- **Average Correct Inclusion (ACI):** Average number of covariates that are correctly included in the model and is given by

$$ACI = \frac{\sum_{i=1}^{M} CI_i}{M} \tag{14}$$

  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.
- **Average False Exclusion (AFE):** Average number of covariates that are excluded from the model although its true effect is non-zero and is given by

$$AFE = \frac{\sum_{i=1}^{M} FE_i}{M} \tag{15}$$

  where FE_i is the number of covariates that are falsely excluded from the model in i^{th} MC replication.
- **Correct Inclusion Ratio (CIR)):** Ratio of replications where all covariates are correctly included in the model to all replications and is given by

$$CIR = \frac{\sum_{i=1}^{M} CII_i}{M} \tag{16}$$

  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.

Performance in parameter estimation is measured using relative bias (RB) and mean squared error (MSE) that are given below,

$$RB = \frac{E(\hat{\beta}) - \beta}{\beta}, MSE = E[(\hat{\beta} - \beta)^2] \tag{17}$$

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

5.1. Simulation study 1: sensitivity analyses

5.1.1. Sensitivity analysis for standard 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 HC hyper-prior which is becoming increasingly popular in the Bayesian literature as a robust alternative [12,23,26,28]. Another one is gamma hyper-prior which is proposed by Park and Casella [25] as a suitable hyper-prior for the tuning parameter. Table 2 lists the hyper-priors considered in this study taking the account for different degrees of informativeness.

Tables 3–6 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 7–9. First of all, the true behavior for variable selection is as follows. For scenarios 1 and 2, ACI should be 3, and for scenario 3 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 2. 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)

Table 3. 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)

Note: NA: No results are available for these parameters.

Table 4. 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)

Note: NA: No results are available for these parameters.

Table 5. 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)

Note: NA: No results are available for these parameters.

Table 6. 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)

Note: NA: No results are available for these parameters.

Table 7. 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

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

Table 8. 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

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

Table 9. 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

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

5.1.2. Sensitivity analysis for EB-GS hyper-prior for λ

In this part, we investigate sensitivity of EB-GS based Bayesian lasso results to hyper-hyper prior settings assumed for r and δ parameters. Since simulations for sensitivity analysis takes too long time, this sensitivity analysis is made by considering just Scenario I. Moreover, we consider that the effects of the coefficients in this scenario can be used to measure the sensitivity of the EB-GS method, in general. 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 10 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 with moderate sample size such as $n = 250$. Table 11 present the performance measures for variable selection. As the sample size increases, all measures get closer to each other. 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, especially for moderate sample size.

Table 10. 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)

Note: NA: No results are available for these parameters.

Table 11. Performances in variable selection under 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)
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

5.2. Simulation study 2: performance of EB-GS

In this section, we proceed with comparisons of performance of EB-GS with hyper-prior distributions used in the previous section. Overall, there are 3 hyper-prior distributions examined per each sample size; (*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 5.1 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 5.

In $Ga(\hat{r}, \hat{\delta})$, $\hat{r}$ and $\hat{\delta}$ are posterior modes that are obtained by using proposed approach presented in Section 3.1.

Table 12. 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)

Note: NA: No results are available for these parameters.

5.2.1. Results in terms of parameter estimations

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 12–17 show RBs and MSEs for this simulation study.

Table 13. 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)

Note: NA: No results are available for these parameters.

Table 14. 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)

Note: NA: No results are available for these parameters.

Table 15. 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)

Note: NA: No results are available for these parameters.

Table 16. 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)

Note: NA: No results are available for these parameters.

In general, RBs and MSEs when EB-GS hyper-prior is used are comparable or lower than those when other hyper-prior distributions are used, especially when $\kappa \geq 2$. This implies that Bayesian lasso with EB-GS hyper-prior for the tuning parameter has improved performance when the final parameter estimations are concerned for linear-circular regression models. Additionally, results show that RBs and MSEs are smaller when $\kappa < 2$ for all methods.

5.2.2. *Results in terms of variable selection*

Tables 18 and 19 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 used are comparable or superior than those when other hyper-prior distributions are used. This shows that proposed strategy appears to supply a clearer distinction between significant and insignificant covariates. With this behavior, the Bayesian lasso with EB-GS hyper-prior support modeling with appropriate dimension including only the important covariates and identifying the best model with increased precision as sample size increases. Additionally, results show that the variable selection measures are superior when $\kappa < 2$ for all methods.

Table 17. 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)	NaN(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)

Note: NA: No results are available for these parameters.

Table 18. 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)
100	ACI	1.850	1.910	1.884	0.514	0.520	0.296	6.960	6.914	6.952
	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
250	ACI	2.805	2.812	2.812	1.403	1.417	1.098	8.000	8.000	8.000
	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

Note: *The true values of variable selection measures.

6. Applications

In this section we apply our method on two different environmental data sets. The first data set (Section 6.1) is the air quality data set commonly used in circular literature. The second data set (Section 6.2) concerns with factors effecting temperature in Sydney where bush fires are regular threat.

Table 19. 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)
100	ACI	2.926	2.918	2.924	1.396	1.426	0.918	7.998	7.998	7.998
	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
250	ACI	3	3	3	2.658	2.686	2.514	8	8	8
	AFE	0	0	0	0.342	0.314	0.486	0	0	0
	CIR	1	1	1	0.686	0.712	0.592	1	1	1

Note: *The true values of variable selection measures.

Bayesian analyses of these data sets are carried out in OpenBUGS. We consider two different chains with 1000000 updates after discarding 10000 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 (18) can be used to calculate RMSEs

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{18}$$

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.

6.1. 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 [5] introduced a method to derive air quality index and constructed a data set consisting of their air quality index and environmental variates such as temperature, wind direction and speed. Johnson and Wehrly [16] 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 [31] 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 20 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

Table 20. 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	

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.

6.2. 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 in the 24 h 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 h. Shapiro–Wilk goodness of fit test [31] 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 6.1. Extensive results are presented in Table 21. 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.

7. 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.

Table 21. 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	

The aim of this article was to initiate a work on Bayesian penalization techniques such as the Bayesian lasso for variable selection in linear-circular regression models to improve the prediction ability of the model and reduce model complexity. In this article, 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. Based on simulations and empirical results, we highlight the following attractive properties.

- 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 important and non-important 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-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.
- The adaptation of Bayesian lasso can straightforwardly be implemented by using a simple Gibbs sampling procedure for linear-circular regression models.

To conclude, 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. Finally, 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.

As a possible avenue for future work, it would be interesting to apply our proposed method for circular-circular or circular-linear regression models. Such an investigation is currently under investigation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Onur Camli http://orcid.org/0000-0003-3885-3781

References

[1] D. Andrews and C. Mallows, *Scale mixtures of normal distributions*, J. R. Stat. Soc. Series B Stat. Methodol. 36 (1974), pp. 99–102.
[2] P. Buhlmann and D.G.S. Van, *Statistics for High-dimensional Data: Methods, Theory and Applications*, Springer, New York, 2011.

[3] O. Camli and Z. Kalaylioglu, *Bayesian predictive model selection in circular random effects models with applications in ecological and environmental studies*, Environ. Ecol. Stat. 28 (2021), pp. 21–34.
[4] J.A Carnicero and M.P. Wiper MP, *A semi-parametric model for circular data based on mixtures of beta distributions*, Stat. Econ. Ser. No. ws081305 (2008), pp. 08–13.
[5] F. De Wiest and F.H. Della, *Suggestions for a realistic definition of an air quality index relative to hydro-carbonaceous matter associated with airborne particles*, Atmos. Environ. (91975), pp. 951–954.
[6] T.D. Downs and K.V. Mardia, *Circular regression*, Biometrika 89 (2002), pp. 683–698.
[7] B. Efron, T. Hastie T, Johnstone I, and R. Tibshirani, *Least angle regression*, Ann. Statist. 32 (2004), pp. 407–499.
[8] Y. Fan and Y.T. Cheng, *Tuning parameter selection in high dimensional penalized likelihood*, J. R. Stat. Soc. Series B Stat. Methodol. 75 (2013), pp. 531–552.
[9] J. Fan and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Statist. Assoc. 96 (2001), pp. 1348–1360.
[10] N.I. Fisher and A.J. Lee, *Regression models for an angular response*, Biometrics 48 (1992), pp. 665–677.
[11] C.J. Flynn, C.M. Hurvich, and J.S. Simonoff, *Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models*, J. Amer. Statist. Assoc. 108 (2013), pp. 1031–1043.
[12] A. Gelman, *Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)*, Bayesian Anal. 1 (2006), pp. 515–534.
[13] P. Hall, E.R. Lee, and B.U. Park, *Bootstrap-based penalty choice for the lasso achieving oracle performance*, Statist. Sinica 19 (2009), pp. 449–471.
[14] C. Hans, *Bayesian lasso regression*, Biometrika 96 (2009), pp. 835–845.
[15] T. Inoue and Y. Nagata, *Study of model evaluation method and of selection method of tuning parameter in Lasso*, Total Qual. Sci. 2 (2016), pp. 27–35.
[16] R.A. Johnson and T.E. Wehrly, *Some angular-linear distributions and related regression*, J. Amer. Statist. Assoc. 73 (1978), pp. 602–606.
[17] S. Kawano, I. Hoshina, K. Shimamura K, and K. Sadanori, *Predictive model selection criteria for Bayesian lasso regression*, J. Jpn. Soc. Comput. Stat. 28 (2015), pp. 67–82.
[18] S. Kim and A. SenGupta, *Inverse circular-linear/linear-circular regression*, Commun. Stat. Theory Methods. 44 (2015), pp. 4772–4782.
[19] M. Kyung, J. Gill, M. Ghosh, and G. Casella, *Penalized regression, standard errors, and Bayesian lassos*, Bayesian Anal. 5 (2010), pp. 369–411.
[20] U. Lund, *Least circular distance regression for directional data*, J. Appl. Stat. 26 (1999), pp. 723–733.
[21] A. Lykou and I. Ntzoufras, *On Bayesian lasso variable selection and the specification of the shrinkage parameter*, Stat. Comput. 23 (2013), pp. 361–390.
[22] H. Mallick and N. Yi, *A new Bayesian lasso*, Stat. Interface. 7 (2014), pp. 571–582.
[23] J. Mulder and L.R. Pericchi, *The matrix-f prior for estimating and testing covariance matrices*, Bayesian Anal. 13 (2018), pp. 1189–1210.
[24] W. Murray, P. Gill, and M. Wright, *Practical Optimization*, Academic Press, New York, 1981.
[25] T. Park and G. Casella, *The Bayesian lasso*, J. Amer. Statist. Assoc. 103 (2008), pp. 681–686.
[26] N.G. Polson and J.G. Scott, *On the half-cauchy prior for a global scale parameter*, Bayesian Anal. 7 (2012), pp. 887–902.
[27] P.K. Ravindran and S.K. Ghosh, *Bayesian analysis of circular data using wrapped distributions*, J. Stat. Theory Pract. 5 (2001), pp. 547–561.
[28] V.E. Sara, L.O. Daniel, and M. Joris, *Shrinkage priors for Bayesian penalized regression*, J. Math. Psychol. 89 (2019), pp. 31–50.
[29] A. SenGupta and S. Bhattacharya, *Finite mixture-based Bayesian analysis of linear-circular models*, Environ. Ecol. Stat. 22 (2015), pp. 667–679.
[30] A. SenGupta and F.I Ugwuowo, *Asymmetric circular-linear multivariate regression models with applications to environmental data*, Environ. Ecol. Stat. 13 (2006), pp. 299–309.

[31] S. Shapiro and M.B. Wilk, *An analysis of variance test for normality (complete samples)*, Biometrika 52 (1965), pp. 591–611.
[32] A.E. Sikaroudi and C. Park, *A mixture of linear-linear regression models for a linear-circular regression*, Stat. Modelling. 21 (2021), pp. 220–243.
[33] P.B. Stephen and A. Gelman, *General methods for monitoring convergence of iterative simulations*, J. Comput. Graph. Stat. 7 (1998), pp. 434–455.
[34] S. Tateishi, H. Matsui, and S. Konishi, *Nonlinear regression modeling via the lasso-type regularization*, J. Stat. Plan. Inference. 140 (2010), pp. 1125–1134.
[35] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Series B Stat. Methodol. 58 (1996), pp. 267–288.
[36] H. Wang, R. Li, and C.L. Tsai, *Tuning parameter selectors for the smoothly clipped absolute deviation method*, Biometrika 94 (2007), pp. 553–568.
[37] H. Wang, B. Li, and C. Leng, *Shrinkage tuning parameter selection with a diverging number of parameters*, J. R. Stat. Soc. Series B Stat. Methodol. 71 (2009), pp. 671–683.
[38] Y. Yu and Y. Feng, *Modified cross-validation for penalized high-dimensional linear regression models*, J. Comput. Graph Stat. 23 (2014), pp. 1009–1027.