Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Bayesian Variable Selection for Beta Regression Model

Ayat Salim Al-Jajawi and Taha Alshaybawee

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq

Emails: ayatjajawi@gmail.com , taha.alshaybawee@qu.edu.iq

ARTICLE INFO

Received: 20 Dec 2024

Revised: 22 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Beta regression has emerged as a widely used technique for modeling the relationship between a response variable and a set of covariates. It assumes that the dependent variable follows a beta distribution, making it especially well-suited for continuous outcomes restricted to the interval (0, 1). This paper presents a Bayesian Lasso framework for variable selection and parameter estimation within the Beta Regression Model (BRM), which is particularly suited for modeling continuous response variables constrained to the (0, 1) interval such as proportions and rates. By incorporating Laplace priors through a hierarchical Bayesian structure, the proposed Bayesian Lasso Beta Regression model enables simultaneous coefficient shrinkage and variable selection, thereby improving both model interpretability and predictive performance. Monte Carlo simulations and a real data analysis are conducted to evaluate and compare the performance of the proposed Bayesian Lasso Beta Regression with the non-Bayesian Beta Regression and Bayesian Beta regression.

Keywords: Bayesian Inference, Beta regression, Lasso, Variable selection, Gibbs sampler.

1. Introduction

The beta regression model has become one of the most popular techniques for modeling the relationship between a response variable and covariates. This model assumes that the dependent variable follows a beta distribution, making it particularly suitable for continuous data constrained within the interval (0,1). The model offers several advantages, including the ability to account for heteroskedasticity and asymmetry in the data, direct interpretability of the regression parameters in relation to the mean of the original variable y, and a more flexible and robust approach for analyzing proportions, rates, and indices constrained within the (0,1) interval. As such, the beta regression model serves as a superior alternative to traditional transformation-based methods. Introduced by (Ferrari and Cribari-Neto 2004) he (0,1) interval providing a framework for handling such data effectively. Beta regression models (BRMs) have been widely applied across various fields. For example, (Erkoç and Sever 1986) used BRMs to analyze body fat percentages, while (Qasim et al. 2021) applied them to assess crude oil proportions after distillation and fractionation. Additionally, (Karlsson and Fraenkel 2020) utilized BRMs to evaluate the color characteristics of hazelnuts.

The maximum likelihood estimator (MLE) is commonly used to estimate model parameters. However, in certain applications, incorporating prior information about the parameters can be beneficial. Such prior information can improve estimation accuracy and is often incorporated as constraints within the model. Beta Regression Models (BRMs) can incorporate both linear and nonlinear constraints, which may be in the form of either equalities or inequalities.(Seifollahi, Bevrani, and Mansson 2024) and (Yang et al. 2022) have explored the use of linear equality constraints in BRMs. Their studies aim to

Research Article

enhance the Beta Maximum Likelihood Estimator (BMLE) and the Beta Liu Estimator (BLE), respectively, by applying shrinkage methods such as the James-Stein, positive James-Stein, and preliminary test methods to constrain the regression parameters. These constraints ensure structural consistency based on physical phenomena or the validity of scientific theories. For example, in applied econometrics, certain coefficient parameters must be non-negative or non-positive (Borders 1989), (Bails and Peppers 1982). Similarly, in hyperspectral imaging, physical considerations require that coefficient parameters remain non-negative (Shaw and Manolakis 2002)

Among penalized regression techniques, the Least Absolute Shrinkage and Selection Operator (LASSO) is one of the most widely used methods in statistical analysis. LASSO applies L1 regularization to the regression model, which encourages sparsity in the coefficients by driving some of them to exactly zero. This makes it particularly effective for variable selection and shrinkage. LASSO is a specific case of the bridge estimator, where the parameter α is set to 1, thereby balancing between model fit and complexity. (Tibshirani 1996) introduced the Least Absolute Shrinkage and Selection Operator (LASSO), which is formulated as the solution to the following optimization problem: minimize the residual sum of squares subject to a constraint on the sum of the absolute values of the regression coefficients. Mathematically, this is expressed as:

$$\min_{\beta}\left(\sum_{i=1}^{n}(y_i - X_i\beta)^2\right) \quad subject\ to\ \lambda\sum_{j=1}^{p}|\beta_j| \le t \quad \dots(1)$$

where λ controls the amount of shrinkage applied to the coefficients, and t is a constant.

Bayesian variable selection using the LASSO (Least Absolute Shrinkage and Selection Operator) is a widely used approach for high-dimensional regression. The Bayesian framework offers a probabilistic alternative to the traditional LASSO, enabling uncertainty quantification in variable selection. The Bayesian LASSO was first introduced by (Park and Casella 2008), who applied Laplace (double-exponential) priors to the regression coefficients, providing a Bayesian interpretation of LASSO. This method enables shrinkage while integrating prior information into the model. In this paper, a Bayesian hierarchical model is construct for estimation and variable selection in the beta regression framework, with modifications and extensions to enhance its performance.

2. Beta Regression

Beta regression is a specialized regression model designed for situations where the dependent variable represents a proportion or percentage constrained within the open interval (0,1). It is particularly useful when the response variable is both bounded and exhibits heteroskedasticity, making conventional linear regression unsuitable.

The beta regression model relies on an alternative parameterization of the beta distribution, where the density function is defined in terms of the mean of the response variable and a precision parameter. The beta density is typically represented as:

$$f(y;p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}y^{p-1}(1-y)^{q-1}\ ,\ \ 0 < y < 1 \dots\dots(2)$$

To construct a regression model for beta-distributed random variables, we begin with the equation that defines the beta distribution density, parameterized by p and q. However, in regression analysis, it is more practical to model the mean of the response variable directly. Additionally, incorporating a precision (or dispersion) parameter into the model is a common approach to account for variability.

To establish a regression framework that incorporates both the mean of the response variable and a precision parameter, we reparameterize the beta density using an alternative formulation. Let: $\mu = p/(p+q)$ and $\gamma = (p+q)$ which $p = \mu\gamma$ and $q = (1-\mu)\gamma$.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Using this parameterization, the expected value and variance of the random variable y are expressed as:

$$E(y) = \mu \qquad var(y) = \frac{v(\mu)}{1+\gamma}$$

$$V(\mu) = \mu(1-\mu)$$

Here, μ represents the mean of the response variable, while γ serves as the precision parameter. For a given μ, larger values of γ lead to a smaller variance $var(y)$, indicating greater concentration of the response variable around its mean. The density function of y under the new parameterization is given by:

$$f(y;\mu,\gamma) = \frac{\Gamma(\gamma)}{\Gamma(\mu\gamma)\Gamma\big((1-\mu)\gamma\big)} y^{\mu\gamma-1}(1-y)^{(1-\mu)\gamma-1}, 0 < y < 1 \ldots..(3)$$

where: $0 < \mu < 1 \quad and \quad \gamma > 0$. This reparameterization enables a regression framework that simultaneously models the mean μ and the precision parameter γ, enhancing the model's flexibility and making it more effective in capturing the variability present in the data.

3. Bayesian Lasso for Beta Regression

The statistical analysis of continuous data constrained within the interval (0,1), such as rates and proportions, requires a probability model that adheres to these boundaries. In this section, we present a Bayesian beta regression modeling framework specifically designed for such data. Bayesian Beta regression is an extension of classical Beta regression that integrates prior distributions for model parameters, enabling probabilistic inference.

Suppose that n independent response variable y_i take values within the interval (0,1) follows a Beta distribution $y_i \sim Beta(\mu_i, \gamma)$where μ_i is the mean of the Beta distribution, often modeled via a link function and γ is the precision parameter, and x_i is a $p \times 1$ vector of covariates. The model allows the response mean to depend on linear predictors by applying the link function $k(.)$ in the following manner:

$$k(\mu_i) = \ln\left(\frac{\mu_i}{1-\mu_i}\right) = \boldsymbol{x}_i'\boldsymbol{\beta}, \quad i = 1, \ldots, n$$

Where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown parameters, the likelihood function can be shown as follows:

$$l(y|\beta,\gamma,\boldsymbol{\mu}) = \prod_{i=1}^{n} \frac{\Gamma(\gamma)}{\Gamma(\mu_i\gamma)\Gamma\big((1-\mu_i)\gamma\big)} y_i^{\mu_i\gamma-1}(1-y_i)^{(1-\mu_i)\gamma-1} \quad \ldots(4)$$

The regression model is typically specified as:

$$g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$$

$$\mu_i = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}$$

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Where $g(.)$ is a link function (e.g., logit, log, or probit), $\boldsymbol{x}_i$ is a vector of covariates. Bayesian Lasso regularization for beta regression can be achieved by assigning a Laplace (double-exponential) prior to the regression coefficients (Park and Casella 2008). Therefore, the distribution of Laplace prior can be expressed as:

$$\pi(\beta_j|\lambda,\sigma) = \frac{\lambda}{2\sigma} e^{\left\{-\frac{\lambda|\beta_j|}{\sigma}\right\}} \quad \dots(5)$$

where λ>0 is the regularization parameter that controls the sparsity of the regression coefficients. To simplify sampling, the Laplace prior is often reformulated as a hierarchical model by expressing it as a scale mixture of normals. This is done by introducing an auxiliary variance-like parameter and assigning it an exponential prior (Andrews and Mallows 1974):

$$\frac{\theta}{2} e^{-\theta|z|} = \int_0^\infty \frac{1}{\sqrt{2\pi s}} \exp\left\{-\frac{z^2}{2s}\right\} \frac{\theta^2}{2} \exp\left\{-\frac{\theta^2 s}{2}\right\} ds \dots(6)$$

Let $\theta = \frac{\lambda}{\sigma}$, then Laplace prior on β can be written as :

$$\prod_{j=1}^{p} \frac{\theta}{2} e^{-\theta|\beta_j|} = \prod_{j=1}^{p} \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} \exp\left\{-\frac{\beta_j^2}{2s_j}\right\} \frac{\theta^2}{2} \exp\left\{-\frac{\theta^2 s_j}{2}\right\} ds_j$$

In this study, Gamma distribution will set as prior to the precision parameter (γ) and θ^2 . The hierarchical model for Bayesian variable selection in Beta regression can be formulated as follows:

$$l(y|\beta,\gamma,\boldsymbol{\mu}) = \prod_{i=1}^{n} \frac{\Gamma(\gamma)}{\Gamma(\mu_i\gamma)\Gamma\big((1-\mu_i)\gamma\big)} y_i^{\mu_i\gamma-1} (1-y_i)^{(1-\mu_i)\gamma-1}$$

$$\pi(\beta,s|\theta^2) = \prod_{j=1}^{p} \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} \exp\left\{-\frac{\beta_j^2}{2s_j}\right\} \frac{\theta^2}{2} \exp\left\{-\frac{\theta^2 s_j}{2}\right\} ds_j \quad \dots(7)$$

$$\pi(\gamma) \sim \gamma^{a-1} \exp(-b\gamma)$$

$$\pi(\theta^2) \sim (\theta^2)^{c-1} \exp(-d\theta^2)$$

Where *a*, *b*, *c* and *d* are the hyperparameters.

4. The Conditional Posterior Distributions:

Based on the hierarchical model (7), the posterior distribution for Bayesian variable selection in beta regression can be constructed as follows:

1- Sample the coefficients $\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{s},\gamma,\theta^2$ from the full conditional posterior distribution of β :

$$\pi(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{s},\gamma,\theta^2) \propto \pi(\boldsymbol{y}\,|\boldsymbol{\beta},\boldsymbol{s},\gamma,\theta^2) \times \pi(\boldsymbol{\beta}|\boldsymbol{s})$$

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$\propto \prod_{i=1}^{n} \frac{\Gamma(\gamma)}{\Gamma(\mu_i\gamma)\Gamma\big((1-\mu_i)\gamma\big)} y_i^{\mu_i\gamma-1}\,(1-y_i)^{(1-\mu_i)\gamma-1} \times \prod_{j=1}^{p} \exp\left\{-\frac{\beta_j^2}{2s_j}\right\}$$

Since the distribution is not common, the Metropolis algorithm will be used to sample $\boldsymbol{\beta}$.

2- Sample $\boldsymbol{s}|\boldsymbol{\beta},\boldsymbol{y},\gamma,\theta^2$ from the following full conditional posterior distribution of $\boldsymbol{s}$:

$$\boldsymbol{\pi}(\boldsymbol{s}|\boldsymbol{\beta},\boldsymbol{y},\gamma,\theta^2) \propto \pi(\boldsymbol{\beta}|\boldsymbol{s}) \times \pi(\boldsymbol{s}|\theta^2)$$

$$\propto \frac{1}{\sqrt{2\pi s_j}} \exp\left\{-\frac{\beta_j^2}{2s_j}\right\} \times \exp\left\{-\frac{\theta^2 s_j}{2}\right\}$$

$$\propto \frac{1}{\sqrt{s_j}} \exp\left\{-\frac{1}{2}\left(\beta_j^2 s_j^{-1} + \theta^2 s_j\right)\right\}$$

The full conditional posterior distribution of $\boldsymbol{s}$ is generalized inverse Gaussian distribution.

3- Sample $\gamma|\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{s},\theta^2$ from the following full conditional posterior distribution of γ:

$$\pi(\gamma|\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{s},\theta^2) \propto \pi(\boldsymbol{y}\,|\boldsymbol{\beta},\boldsymbol{s},\gamma,\theta^2) \times \pi(\gamma)$$

$$\propto \prod_{i=1}^{n} \frac{\Gamma(\gamma)}{\Gamma(\mu_i\gamma)\Gamma\big((1-\mu_i)\gamma\big)} y_i^{\mu_i\gamma-1}\,(1-y_i)^{(1-\mu_i)\gamma-1} \times \gamma^{a-1}\exp(-b\gamma)$$

Since the distribution is uncommon, the Metropolis algorithm will be used to sample γ.

4- Sample $\theta^2|\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{s},\gamma$ from the following full conditional posterior distribution of θ^2:

$$\pi(\theta^2|\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{s},\gamma) \propto \pi(\boldsymbol{s}\,|\theta^2) \times \pi(\theta^2)$$

$$\propto \prod_{j=1}^{p} \frac{\theta^2}{2} \exp\left\{-\frac{\theta^2 s_j}{2}\right\} \times (\theta^2)^{c-1}\exp(-d\theta^2)$$

$$\propto (\theta^2)^{p+c-1} exp\left\{-\theta^2\left(\sum_{j=1}^{p}\frac{s_j}{2} + d\right)\right\}$$

The full conditional posterior distribution of θ^2 is a Gamma distribution.

5. Simulation studies:

In this section, we conduct Monte Carlo simulations to evaluate the performance of Bayesian regularized Beta regression, comparing it with the non-Bayesian Beta approach proposed by (Ferrari and Cribari-Neto 2004) and the Bayesian Beta regression method proposed by (Kottas and Gelfand 2001). In this study, we propose performance criteria such as the bias and standard deviation of the

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

estimated parameters, as well as the mean squared error and mean absolute error of the model. We consider three simulation scenarios for the parameter coefficients, which are similar to those presented in (Zhu et al. 2008), these scenarios as follows:

- Scenario one: β = (3, 1.5, 0, 0, 2, 0, 0, 0), representing a sparse case.
- Scenario two: β = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85), representing a dense case.
- Scenario three: β = (5, 0, 0, 0, 0, 0, 0, 0), representing a very sparse case.

The explanatory variables $\boldsymbol{x}'$ are drawn from a multivariate standard normal distribution $N(0, \Sigma)$ where the *(i, j)*th entry of Σ is defined as $0.5^{|i-j|}$. The data is generated for two different sample sizes: n=50 and 150. The number of explanatory variables is set to p=8, and the precision parameter is set to $\gamma = 4$. The response variable will be generated as:

$$y_i \sim Beta(\mu_i\gamma, (1-\mu_i)\gamma)$$

where μ_i represents the mean response for the *i-th* observation, and γ is a precision parameter that controls the dispersion of the Beta distribution. For Bayesian inference at each quantile, a total of 12,000 iterations will be run, with the first 2,000 iterations discarded as burn-in. For 100 replications of our experiment, the standard deviation and bias are reported in the following tables and figures.

Table (1): show the standard deviation and bias to the first scenario at each sample size.

Sample size	Methods	criteria	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\gamma}$
50	BReg	SD	1.2047	0.4637	0.5351	0.4209	0.7264	0.4162	0.2661	0.3779	0.9372
		Bias	2.3207	0.9722	0.1988	0.2049	1.2874	0.1176	0.1547	0.0938	0.8676
	BBReg	SD	0.3283	0.2339	0.1659	0.3154	0.2420	0.3984	0.3677	0.2263	0.5162
		Bias	0.6240	0.3146	0.0029	0.0629	0.3122	0.2342	0.0044	0.1033	2.9488
	BLBReg	SD	0.2787	0.1733	0.1287	0.1495	0.1522	0.2844	0.2372	0.1514	0.3705
		Bias	0.1099	0.0576	0.0340	0.0554	0.1030	0.1027	0.0453	0.1152	0.2659
150	BReg	SD	0.2268	0.3413	0.3236	0.4151	0.3318	0.3172	0.3114	0.3843	0.6666
		Bias	2.4632	1.3225	0.1246	0.2178	1.6188	0.1074	0.1547	0.1113	0.8417
	BBReg	SD	0.3002	0.2423	0.3268	0.2983	0.1699	0.3550	0.2620	0.2929	0.0042
		Bias	1.5442	0.7932	0.0632	0.0402	0.9798	0.0424	0.0638	0.0959	2.7632
	BLBReg	SD	0.2343	0.1578	0.1599	0.1911	0.1371	0.1628	0.1409	0.1732	0.2901
		Bias	0.1005	0.0505	0.0196	0.0169	0.0473	0.0077	0.0740	0.0512	0.1509

Table (1) presents a comparative analysis of three estimation methods Classical Beta Regression (BReg), Bayesian Beta Regression (BBReg), and Bayesian Lasso Beta Regression (BLBReg) under the first simulation scenario, which represents a sparse setting where only a subset of the regression coefficients is non-zero. The evaluation focuses on the standard deviation (SD) and bias of the estimated regression coefficients($\hat{\beta}_1$ *to* $\hat{\beta}_8$) and the precision parameter ($\hat{\gamma}$), across two sample

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

sizes: n = 50 and n = 150. Across both sample sizes, the BLBReg method consistently delivers superior performance, achieving the lowest bias and standard deviation for nearly all parameters. This demonstrates its robustness and accuracy in sparse settings, where only a few predictors are relevant. By effectively shrinking the coefficients of irrelevant variables toward zero, BLBReg enhances model parsimony and reduces estimation error.

In contrast, BBReg exhibits comparatively higher bias, particularly in estimating the precision parameter$(\hat{\gamma})$, with its performance declining further in smaller samples. Additionally, BBReg shows greater variability across several coefficient estimates, indicating reduced stability in sparse scenarios.

The BReg approach, while displaying acceptable variability for some coefficients, generally suffers from higher bias most notably for larger coefficients such as $\hat{\beta}_1$ and $\hat{\beta}_5$. This underscores the limitations of classical methods in handling sparse, high-noise data environments.Overall, these results highlight the advantages of the Bayesian Lasso Beta Regression approach in sparse modeling contexts, offering improved accuracy, stability, and variable selection capability even when sample sizes are limited.

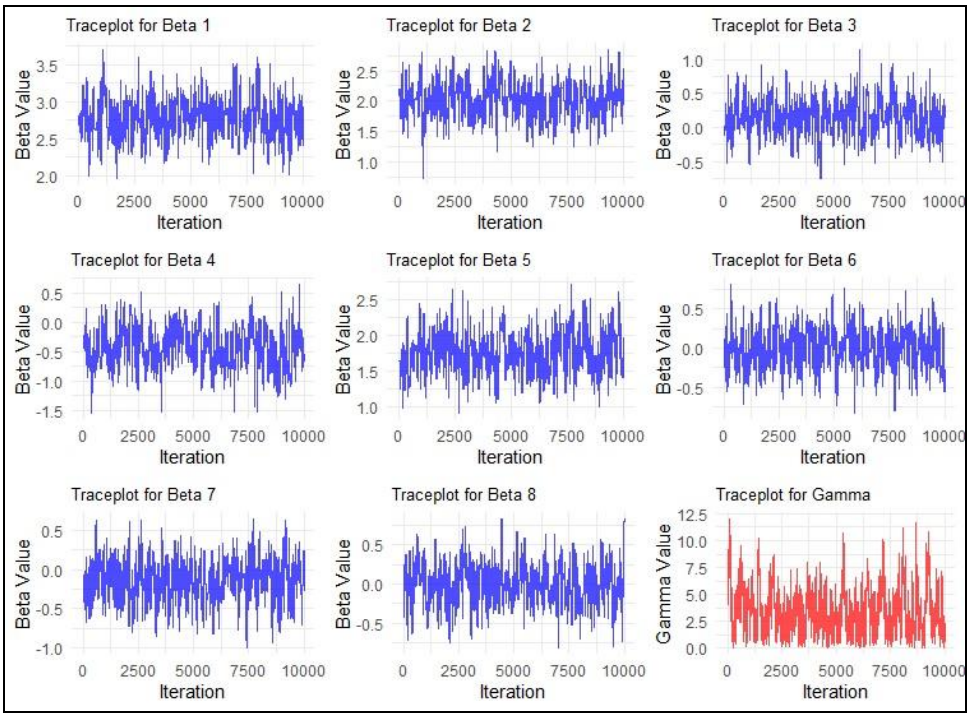


Figure (1): shows the trace plots for the estimated parameters and the precision parameter to the first scenario.

Figure (1) illustrates the superior performance of the Bayesian Lasso Beta Regression (BLBReg) model in the first simulation scenario, which is defined by a sparse structure with few active predictors. Among the three models compared, BLBReg achieves the lowest levels of bias and standard deviation in estimating the regression coefficients. These results underscore the model's effectiveness in both accurate parameter estimation and efficient variable selection, particularly in sparse data settings.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

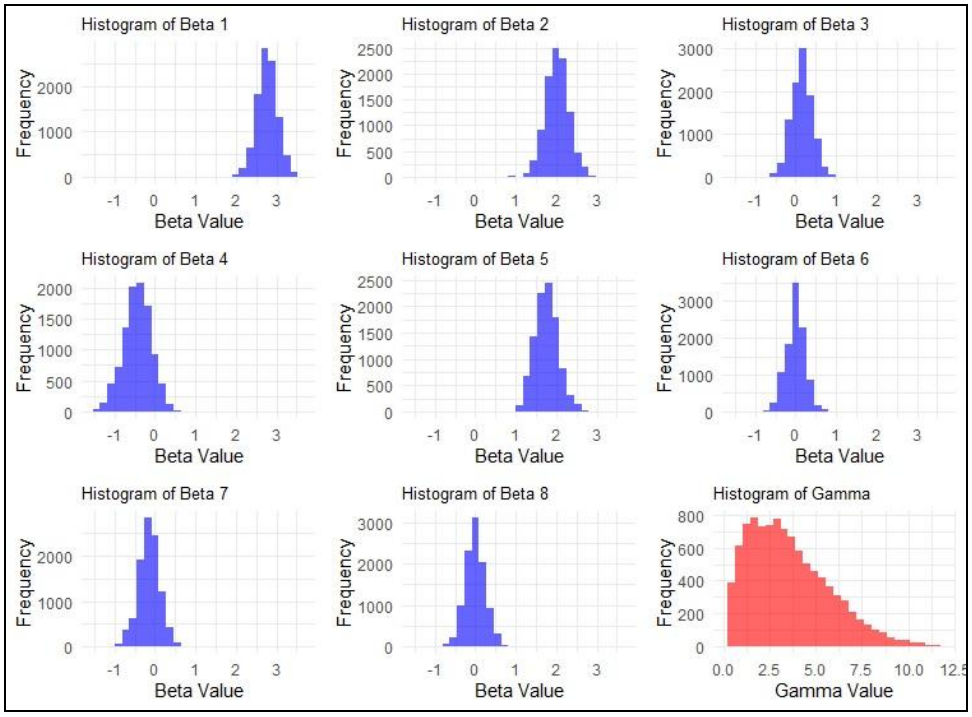


Figure (2): shows the histogram for the estimated parameters and the precision parameter to the first scenario.

Figure (2) displays the posterior distributions of the regression coefficients ($\hat{\beta}_1 \, to \, \hat{\beta}_8$) and the precision parameter ($\hat{\gamma}$) obtained from the Bayesian Lasso Beta Regression (BLBReg) model. The distributions of $\hat{\beta}_1, \hat{\beta}_2$, and $\hat{\beta}_5$ are distinctly centered away from zero, indicating their substantial contribution to explaining the response variable. In contrast, the remaining coefficients are concentrated near zero, suggesting minimal or no influence on the outcome. These findings demonstrate the effectiveness of the Bayesian Lasso framework in performing variable selection by shrinking non-informative coefficients. Furthermore, the posterior distribution of the precision parameter($\hat{\gamma}$) exhibits a right-skewed pattern, with most values falling between 2 and 5, indicating a moderate to high degree of precision in the model's estimates.

Table (2): show the standard deviation and bias to the second scenario at each sample size.

Sample size	Methods	criteria	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\gamma}$
50	BReg	SD	0.1213	0.1678	0.3833	0.3147	0.3150	0.2110	0.2754	0.2494	0.7575
		Bias	0.1640	0.0424	0.0411	0.0355	0.1080	0.1279	0.1515	0.0167	1.2203
	BBReg	SD	0.3413	0.2260	0.5711	0.4586	0.3562	0.3602	0.3975	0.3362	0.1185
		Bias	0.5169	0.4040	0.3202	0.4270	0.5616	0.2362	0.6068	0.3130	2.9485
	BLBReg	SD	0.1423	0.1417	0.3151	0.2924	0.2759	0.1909	0.2218	0.2517	0.0857
		Bias	0.1031	0.0259	0.0941	0.1059	0.0111	0.1010	0.0482	0.0523	0.0131
150	BReg	SD	0.1492	0.2214	0.4202	0.1296	0.3708	0.2835	0.3059	0.3116	0.4818

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	Bias	0.0757	0.0814	0.2019	0.0606	0.2344	0.0751	0.1411	0.1689	0.3255
BBReg	SD	0.2825	0.2848	0.3061	0.2117	0.2083	0.4063	0.2756	0.1963	0.1103
	Bias	0.5857	0.6774	0.6214	0.7377	0.4631	0.6442	0.6414	0.5741	2.9839
BLBReg	SD	0.1290	0.1317	0.1365	0.0734	0.0767	0.2317	0.1461	0.0941	0.1214
	Bias	0.0031	0.0144	0.0003	0.0370	0.0375	0.0055	0.0004	0.0223	0.1567

Table (2) provides a comparative evaluation of three estimation under the second simulation scenario, which represents a dense case where all regression coefficients are equal ($i.e., \beta = 0.85$). The performance of each method is assessed using standard deviation (SD) and bias for the estimated coefficients ($\hat{\beta}_1 to \hat{\beta}_8$) and the precision parameter ($\hat{\gamma}$), across two sample sizes (n = 50 and n = 150).

In both sample sizes, the BLBReg model consistently outperforms the other methods, achieving the lowest bias and standard deviation for nearly all coefficients. By contrast, BBReg exhibits relatively high bias particularly in estimating the precision parameter and greater variability, especially in smaller samples. While BReg performs reasonably well, it falls short of the precision and accuracy achieved by BLBReg.

These results affirm the effectiveness of the Bayesian Lasso approach in high-dimensional modeling contexts, particularly its ability to select and estimate significant predictors with greater reliability while minimizing estimation error.

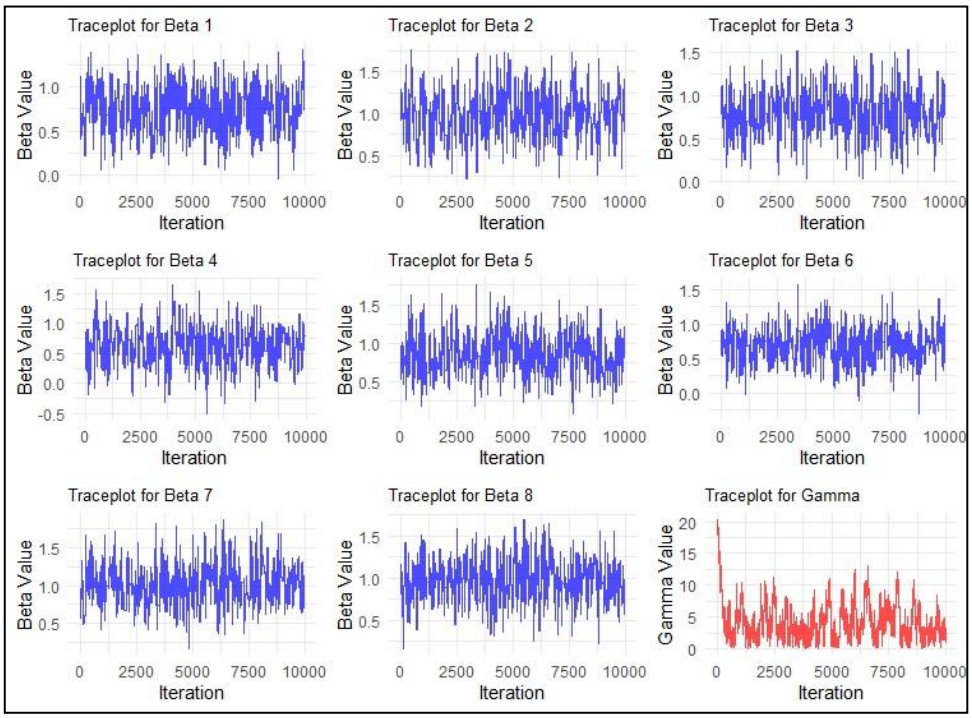


Figure (3): shows the trace plots for the estimated parameters and the precision parameter to the second scenario.

Figure (3) illustrates the posterior distributions of the regression coefficients ($\hat{\beta}_1 to \hat{\beta}_8$) and the precision parameter ($\hat{\gamma}$) obtained from the Bayesian Lasso Beta Regression (BLBReg) model under

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

the second simulation scenario, representing a dense setting in which all covariates are equally relevant.

The figure reveals that all regression coefficients are distinctly centered away from zero, confirming the model's effectiveness in identifying the significance of all predictors. This behavior is expected in dense scenarios, where no covariate should be excluded. Furthermore, the posterior distributions are narrow and smooth, indicating low variability and high estimation precision. These findings align with the numerical results reported in Table (2), underscoring the model's capacity to generate stable and accurate coefficient estimates.

Moreover, the posterior distribution of the precision parametercentered and exhibits -is well $\hat{\gamma}$) (minimal dispersion. This reflects a high degree of confidence in the model's ability to account for variability in the data, reinforcing its robustness and reliability in dense modeling scenarios.

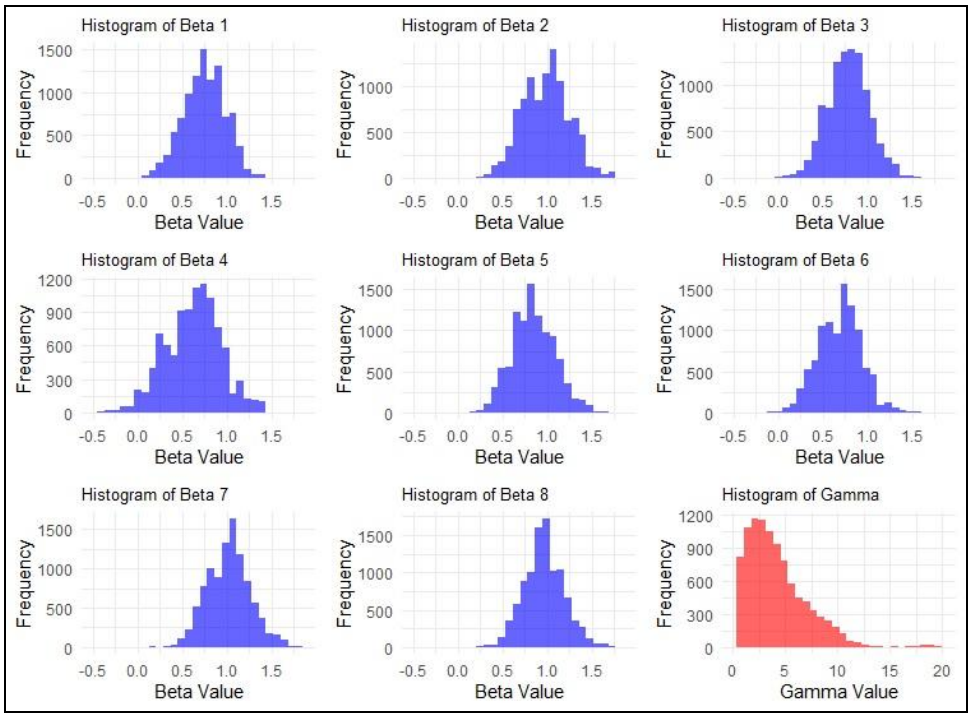


Figure (4): shows the histogram for the estimated parameters and the precision parameter to the second scenario.

Figure (4) displays the posterior distributions of the regression coefficients ($\hat{\beta}_1 \, to \, \hat{\beta}_8$) and the precision parameter ($\hat{\gamma}$) obtained from the Bayesian Lasso Beta Regression (BLBReg) model under the dense simulation scenario with an increased sample size of n = 150.

As anticipated in a dense setting, the distributions of all regression coefficients are distinctly centered away from zero, reaffirming the model's ability to accurately identify and retain all relevant predictors. Compared to the results in Figure (3), which is based on a smaller sample size, the coefficient distributions in Figure (4) appear sharper and more concentrated. This reflects a notable improvement in estimation precision and reduced variability attributable to the larger sample size.

Similarly, the posterior distribution of the precision parameter γ is more tightly concentrated and exhibits lower dispersion, indicating increased certainty in the model's estimation of overall variability. These visual results are consistent with the quantitative outcomes reported in Table (2),

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

which demonstrate reductions in both bias and standard deviation, further validating the model's robustness and reliability in higher-sample dense scenarios.

Table (3): show the standard deviation and bias to the third scenario at each sample size.

Sample size	Methods	criteria	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\gamma}$
50	BReg	SD	0.2711	0.2517	0.3475	0.2421	0.3273	0.3931	0.3246	0.2596	1.5575
		Bias	4.2665	0.0114	0.0669	0.0555	0.2175	0.0188	0.0909	0.0404	1.1686
	BBReg	SD	0.6184	0.3406	0.2198	0.3689	0.5404	0.3926	0.4314	0.2972	0.8091
		Bias	0.4984	0.0066	0.0160	0.2115	0.1558	0.0733	0.2961	0.1990	2.9685
	BLBReg	SD	1.3554	0.1554	0.3261	0.1562	0.2734	0.2146	0.2242	0.3631	0.6874
		Bias	0.2214	0.0721	0.1215	0.0160	0.1450	0.0182	0.0350	0.0243	0.1960
150	BReg	SD	0.3964	0.1411	0.1981	0.3808	0.1428	0.3244	0.2452	0.2030	1.0434
		Bias	4.6766	0.0646	0.0175	0.1749	0.0409	0.0611	0.0157	0.0689	0.4506
	BBReg	SD	0.2138	0.2481	0.2644	0.1964	0.3672	0.2710	0.2358	0.1986	0.9023
		Bias	2.3044	0.1285	0.0529	0.0387	0.0292	0.1074	0.0380	0.0854	2.9884
	BLBReg	SD	0.2482	0.1198	0.1327	0.1681	0.1955	0.1534	0.1726	0.0859	0.6325
		Bias	0.0027	0.0034	0.0374	0.0260	0.0246	0.0857	0.0053	0.0001	0.0887

Table (3) summarizes the bias and standard deviation of the estimated parameters under a highly sparse scenario, in which only $\hat{\beta}_1$ is non-zero. Across both sample sizes (n = 50 and n = 150), the BLBReg model exhibits superior performance, achieving the lowest bias for the active coefficient ($\hat{\beta}_1$) while successfully shrinking the remaining coefficients ($\hat{\beta}_2$ $through$ $\hat{\beta}_8$) toward zero.

It is important to highlight that for n = 50, the standard deviation of β̂1 in BLBReg is relatively large (1.3554). This elevated variability is expected due to the combination of a high true coefficient value $(\hat{\beta}_1 = 5)$ and a limited sample size, both of which contribute to increased uncertainty in the posterior distribution. Nevertheless, the model maintains high estimation accuracy and strong variable selection capacity.

By contrast, BReg and BBReg display greater bias and variability in most coefficients, particularly in estimating the precision parameter $(\hat{\gamma})$. Additionally, both methods show limited effectiveness in distinguishing between relevant and irrelevant predictors, resulting in less sparse and more unstable estimates.

In summary, BLBReg stands out as the most dependable approach under sparse conditions, providing accurate parameter estimates and robust variable selection, even when data are limited.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

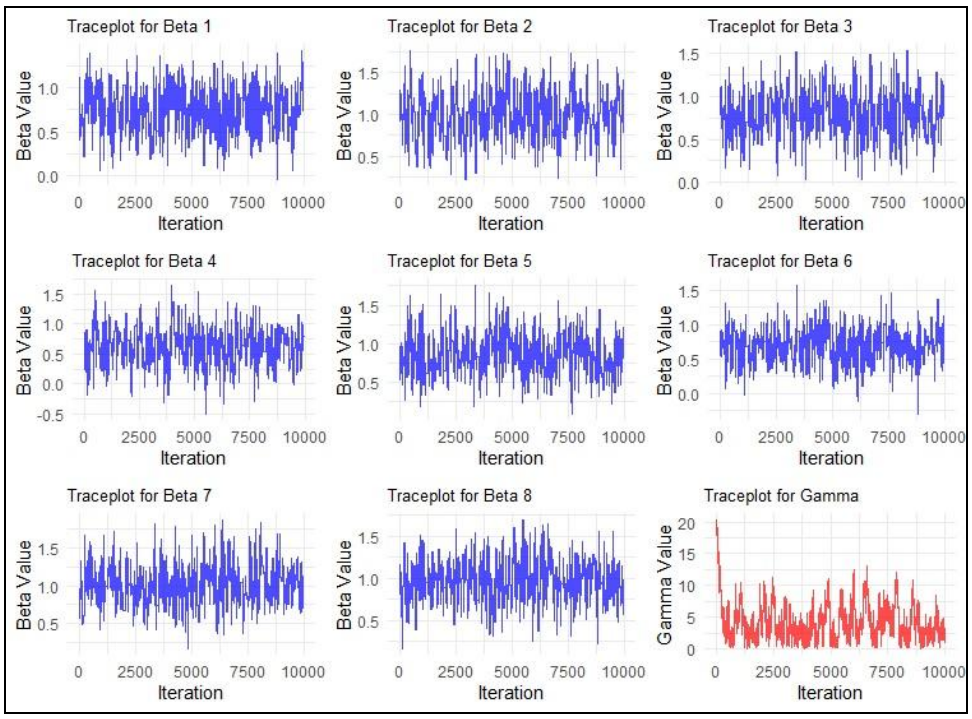


Figure (5): shows the trace plots for the estimated parameters and the precision parameter to the third scenario.

Figure (5) displays the posterior distributions of the regression coefficients ($\hat{\beta}_1 to\, \hat{\beta}_8$) and the precision parameter $(\hat{\gamma})$ under the third simulation scenario, characterized by a very sparse structure, using the Bayesian Lasso Beta Regression (BLBReg) model with a sample size of n = 50.

The distribution of $\hat{\beta}_1$ is clearly centered away from zero, indicating its importance and confirming that the model accurately identifies the true active variable. In contrast, the posterior distributions of are tightly clustered around zero, demonstrating the model's effectiveness in shrinking $\hat{\beta}_8$ through$\hat{\beta}_2$ irrelevant coefficients and enforcing sparsity.

This pattern highlights the strong variable selection capability of the BLBReg model, even in scenarios with limited data and a high proportion of non-informative predictors. Furthermore, the posterior distribution of the precision parameter$(\hat{\gamma})$ is moderately concentrated, indicating that the model maintains reasonable certainty in estimating overall variability despite the small sample size.

These results underscore the robustness and reliability of the Bayesian Lasso approach in sparse settings with limited observations.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

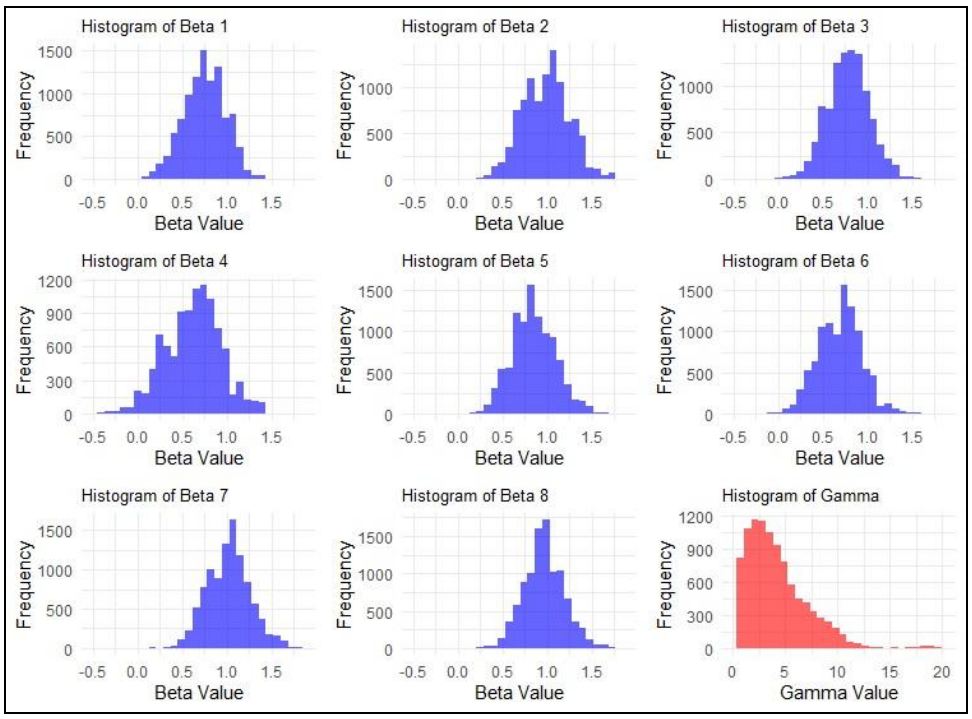


Figure (6): shows the histogram for the estimated parameters and the precision parameter to the third scenario.

Figure (6) presents the posterior distributions of the regression coefficients ($\hat{\beta}_1 \, to \, \hat{\beta}_8$) and the precision parameter$(\hat{\gamma})$ under the third simulation scenario characterized by a very sparse structure using the Bayesian Lasso Beta Regression (BLBReg) model with an increased sample size of n = 50.

The posterior distribution of $\hat{\beta}_1$ remains sharply centered away from zero, reaffirming its significance in the model. In contrast, the distributions of $\hat{\beta}_2$ through $\hat{\beta}_8$ are narrowly concentrated around zero, indicating that the model continues to effectively suppress non-informative variables as the sample size increases.

Moreover, the posterior distribution of the precision parameter $(\hat{\gamma})$becomes noticeably more concentrated compared to the smaller sample case, reflecting increased certainty and enhanced stability in the model's estimation process. These outcomes further emphasize the robustness and efficiency of the BLBReg model in handling sparse scenarios, particularly when supported by larger datasets.

Table (4): show the MSE and MAE for all scenarios at each sample size

Model	Sample size	criteria	Breg	BBReg	BLBReg
Model 1	50	MSE	0.1246	0.1236	0.0896
		MAE	0.1676	0.1580	0.1040
	150	MSE	0.1252	0.1274	0.0836
		MAE	0.1751	0.1866	0.1102
Mode 2	50	MSE	0.2142	0.1126	0.1050

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

		MAE	0.2461	0.1707	0.1255
	150	MSE	0.2185	0.1320	0.1198
		MAE	0.2290	0.1810	0.1344
Model 3	50	MSE	0.1328	0.1348	0.0833
		MAE	0.1436	0.1435	0.1065
	150	MSE	0.1350	0.1421	0.0921
		MAE	0.1491	0.1495	0.1160

Table (4) presents a comparative analysis of the predictive performance of the three models Classical Beta Regression (BReg), Bayesian Beta Regression (BBReg), and Bayesian Lasso Beta Regression (BLBReg) across the three simulation scenarios. Mean Squared Error (MSE) and Mean Absolute Error (MAE) are used as evaluation metrics for two sample sizes (n = 50 and n = 150).

The results clearly indicate that BLBReg consistently achieves superior predictive accuracy, yielding the lowest MSE and MAE values across all scenarios and sample sizes. The model exhibits particularly strong performance in sparse and very sparse scenarios (Model 1 and Model 3), where variable selection is critical. Even in the dense setting (Model 2), BLBReg maintains competitive performance, demonstrating its adaptability and effectiveness under varying degrees of data sparsity.

These findings confirm the robustness and generalizability of the BLBReg model in delivering accurate predictions across a range of modeling contexts.

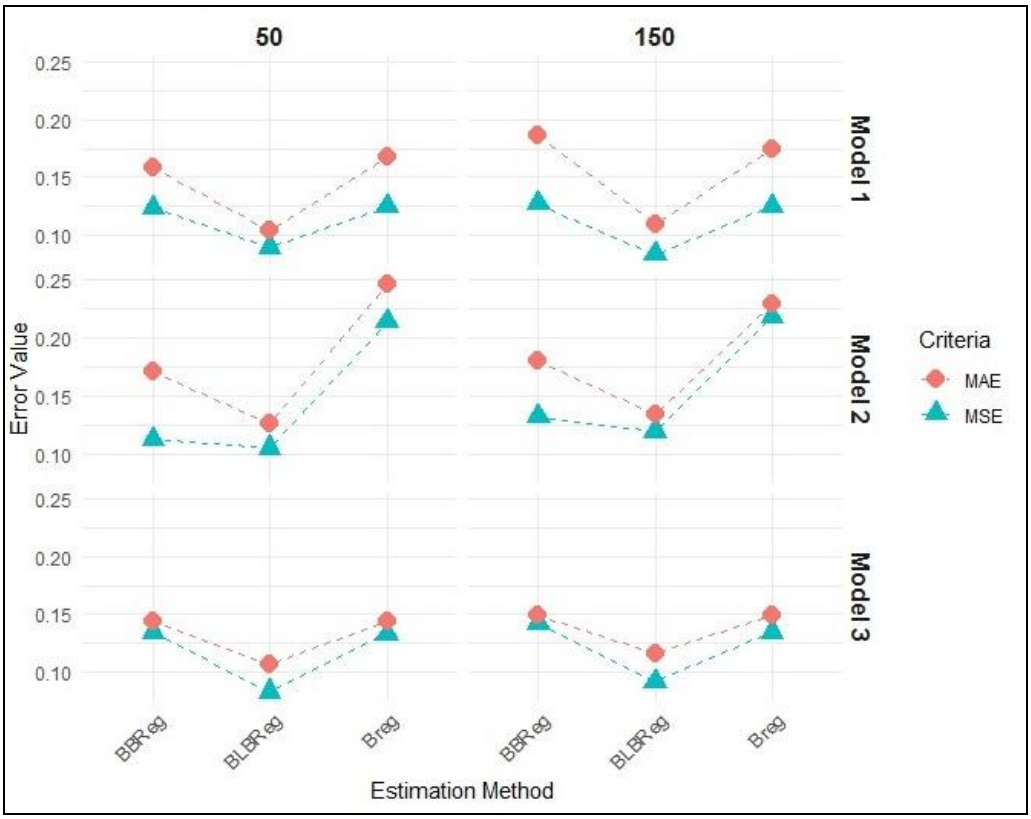


Figure (7): comparison of MSE and MAE for different scenarios and sample sizes.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Figure (7) provides a visual comparison of the predictive accuracy of the three regression models BReg, BBReg, and BLBReg across the three simulation scenarios, using Mean Squared Error (MSE) and Mean Absolute Error (MAE) as evaluation metrics for both sample sizes (n = 50 and n = 150).

The figure clearly illustrates that BLBReg consistently outperforms the competing methods, achieving the lowest MSE and MAE values across all scenarios and sample sizes. The performance gains are especially pronounced in the sparse and very sparse settings (Model 1 and Model 3), where effective variable selection is crucial. Even in the dense scenario (Model 2), BLBReg maintains highly competitive performance, often matching or exceeding that of BBReg.

These graphical findings corroborate the numerical results reported in Table (4), further reinforcing the strength, flexibility, and predictive reliability of the Bayesian Lasso Beta Regression model under diverse data conditions.

5. Real Data Application

The Gasoline Yield dataset, initially compiled by Prater (1956), captures the proportion of crude oil converted into gasoline through distillation and fractionation processes. In a later analysis, Atkinson (1985) applied a linear regression model to this dataset and observed notable asymmetry in the residuals, indicating the presence of both unusually large and small prediction errors.

In the present study, a controlled level of data contamination was introduced by systematically altering the initial values of the explanatory variables by 10%. This modification was implemented to assess the robustness and stability of the regression models when exposed to mild distortions in the input data.

Dataset Description

The Gasoline Yield dataset contains **32 observations** of **6 variables** related to gasoline reduction:

Variable	Description
yield	Proportion of crude oil converted into gasoline (response variable)
gravity	API gravity of crude oil
pressure	Vapor pressure of crude oil
temp10	Temperature at which 10% of crude oil has vaporized
temp	Temperature at which 50% of crude oil has vaporized
temp90	Temperature at which 90% of crude oil has vaporized

Table (5): shows the estimated parameters and the precision parameter to the real data.

Methods	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$
BReg	-0.02636	0.10792	-0.21780	0.74511	-0.20016	25.01168
BBReg	0.03681	0.21296	-0.15650	0.94126	-0.30791	0.54139
BLBReg	0.00000	0.11956	-0.21898	0.60211	-0.08519	4.40813

Table (5) presents the estimated regression coefficients and the precision parameter ($\hat{\gamma}$) for the BReg, BBReg, and BLBReg models, based on the analysis of the Gasoline Yield dataset.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The results clearly demonstrate that the BLBReg model provides more stable and realistic parameter estimates compared to its counterparts. Notably, BLBReg yields a well-balanced estimate of the precision parameter ($\hat{\gamma} = 4.41$), in contrast to the excessively high value obtained by BReg and the significantly underestimated value produced by BBReg.

In addition, BLBReg successfully shrinks the coefficients associated with non-informative predictors toward zero while preserving those linked to meaningful covariates. This highlights the model's ability to perform effective variable selection and maintain robustness in the presence of mild data contamination, making it a reliable tool for real-world regression applications involving bounded response variables.

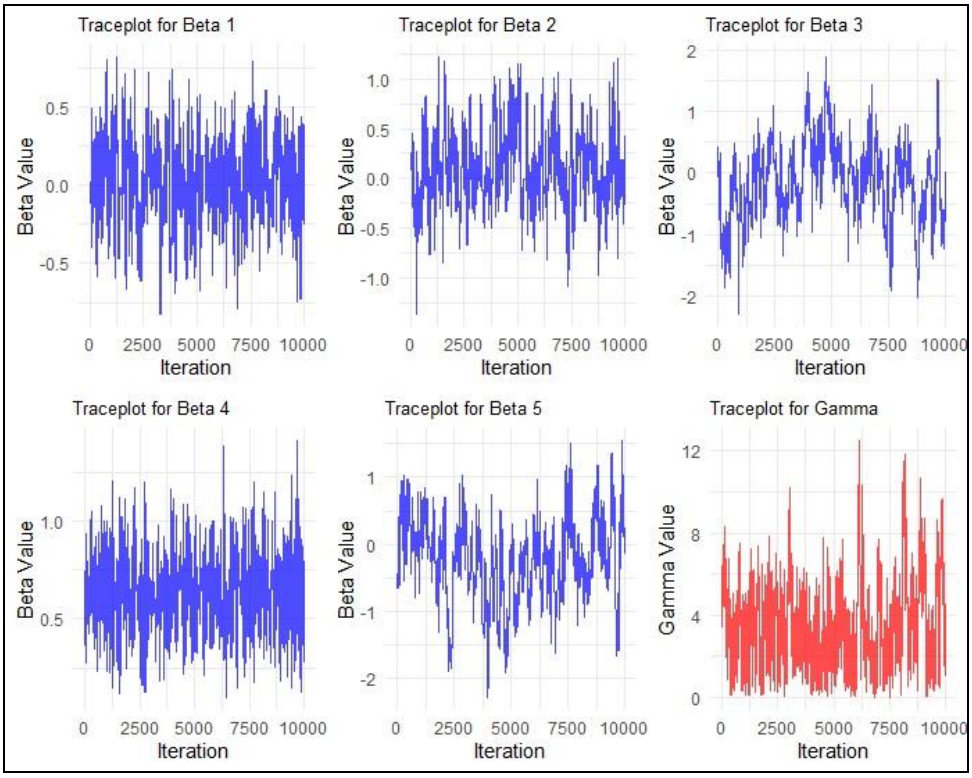


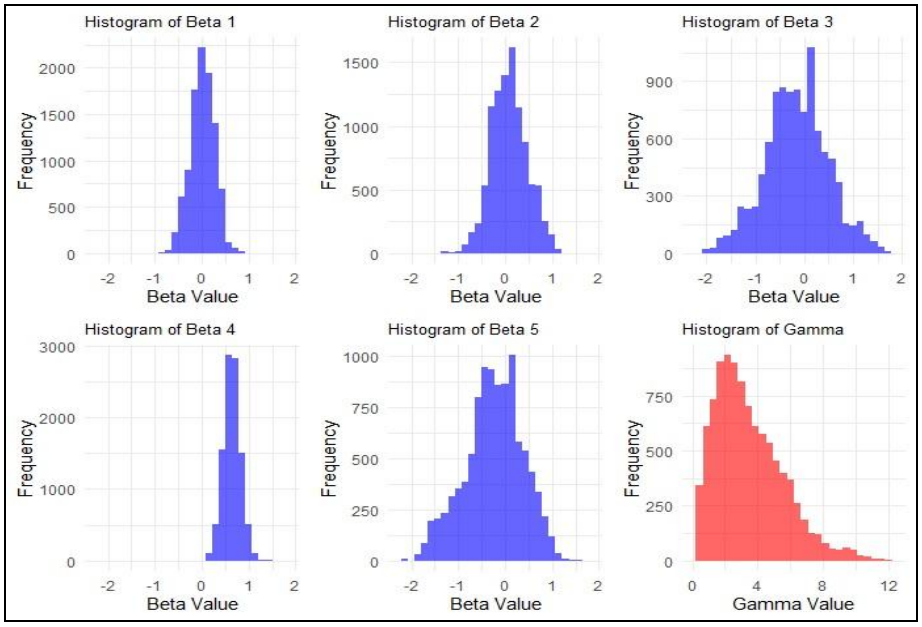
Figure (8): shows the trace plots for the estimated parameters and the precision parameter to the real data

Figure (8) demonstrates the posterior distributions of the estimated regression coefficients and the precision parameter $(\hat{\gamma})$for the BLBReg model based on the Gasoline Yield dataset. The figure shows that certain coefficients such as $\hat{\beta}_2$ and $\hat{\beta}_4$ are clearly centered away from zero, suggesting their significance in explaining the response variable.

In contrast, other coefficients are either concentrated around zero or heavily shrunk, indicating that the model has effectively reduced the influence of irrelevant predictors. The posterior distribution of $(\hat{\gamma})$ also appears well-behaved and moderately concentrated, reflecting a stable and precise estimation of the model's variability.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



Figure(9): shows the histogram for the estimated parameters and the precision parameter to the real data

Figure (9) provides a graphical comparison of the predictive performance of the BReg, BBReg, and BLBReg models on the Gasoline Yield dataset, using Mean Squared Error (MSE) and Mean Absolute Error (MAE) as evaluation metrics. The figure clearly demonstrates that the BLBReg model attains the lowest MSE and MAE values, indicating superior prediction accuracy compared to both BReg and BBReg.

This enhanced performance underscores the robustness and flexibility of the BLBReg model, particularly in scenarios involving mild data contamination. Furthermore, the visual results align closely with the numerical findings presented in Table (6), reinforcing the conclusion that BLBReg is the most accurate and reliable choice for regression modeling when the response variable is constrained within the (0,1) interval.

Table (6): show the MSE and MAE to the real data

Methods	BReg	BBReg	BLBReg
MSE	0.09566	0.10056	0.08325
MAE	0.30308	0.31831	0.27029

Table (6) summarizes the predictive performance of the BReg, BBReg, and BLBReg models on the Gasoline Yield dataset, evaluated using Mean Squared Error (MSE) and Mean Absolute Error (MAE). The results clearly indicate that BLBReg delivers the most accurate predictions, achieving the lowest values for both error metrics.

While BReg and BBReg provide reasonable performance, they are consistently outperformed by BLBReg, particularly in the presence of mild data contamination. These findings further support the effectiveness and robustness of the Bayesian Lasso Beta Regression model for analyzing bounded response data, especially in applications where high predictive accuracy is essential.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6. Conclusions

This study introduces a novel and effective Bayesian Lasso Beta Regression (BLBReg) model, tailored for analyzing continuous response variables bounded within the (0,1) interval, such as rates and proportions. By integrating the Bayesian Lasso framework into the beta regression context, the proposed model effectively addresses key challenges in variable selection, parameter shrinkage, and predictive stability particularly in high-dimensional settings characterized by sparsity in the covariate space. The hierarchical Bayesian formulation employing Laplace priors expressed as scale mixtures of normal enables efficient posterior sampling through Gibbs sampling and the Metropolis algorithm, while also allowing the incorporation of prior information into the model structure.

Simulation results across multiple scenarios with varying levels of sparsity and different sample sizes consistently demonstrate the superiority of BLBReg over classical Beta Regression (BReg) and standard Bayesian Beta Regression (BBReg). Specifically, BLBReg yields lower bias and standard deviation, improved estimation accuracy, and enhanced model sparsity by correctly identifying and retaining only the influential predictors. These advantages are further supported by reduced Mean Squared Error (MSE) and Mean Absolute Error (MAE), confirming the model's reliability and accuracy in prediction.

Moreover, BLBReg demonstrates strong robustness when applied to real-world data, as shown in its performance on the Gasoline Yield dataset under controlled contamination. Unlike other methods, it maintains interpretability and estimation stability by shrinking irrelevant coefficients and producing well-behaved posterior distributions for both the regression coefficients and the precision parameter$(\hat{\gamma})$.

7. References

[1] Andrews, David F., and Colin L. Mallows. 1974. "Scale Mixtures of Normal Distributions." *Journal of the Royal Statistical Society: Series B (Methodological)* 36(1):99–102.

[2] Bails, Dale G., and Larry C. Peppers. 1982. *Business Fluctuations: Forecasting Techniques and Applications*. Prentice-Hall New Jersey.

[3] Borders, Bruce E. 1989. "Systems of Equations in Forest Stand Modeling." *Forest Science* 35(2):548–56.

[4] Erkoç, Sß, and R. Sever. 1986. "1/N Expansion for a Mie-Type Potential." *Physical Review D* 33(2):588.

[5] Ferrari, Silvia, and Francisco Cribari-Neto. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31(7):799–815.

[6] Karlsson, Ulf, and Carl-Johan Fraenkel. 2020. "Covid-19: Risks to Healthcare Workers and Their Families." *Bmj* 371.

[7] Kottas, Athanasios, and Alan E. Gelfand. 2001. "Bayesian Semiparametric Median Regression Modeling." *Journal of the American Statistical Association* 96(456):1458–68.

[8] Park, Trevor, and George Casella. 2008. "The Bayesian Lasso." *Journal of the American Statistical Association* 103(482):681–86.

[9] Qasim, Waqas, Longlong Xia, Shan Lin, Li Wan, Yiming Zhao, and Klaus Butterbach-Bahl. 2021. "Global Greenhouse Vegetable Production Systems Are Hotspots of Soil N2O Emissions and Nitrogen Leaching: A Meta-Analysis." *Environmental Pollution* 272:116372.

[10] Seifollahi, Solmaz, Hossein Bevrani, and Kristofer Mansson. 2024. "Bayesian Analysis of the Beta Regression Model Subject to Linear Inequality Restrictions with Application." *ArXiv Preprint ArXiv:2401.13787*.

[11] Shaw, Gary, and Dimitris Manolakis. 2002. “Signal Processing for Hyperspectral Image Exploitation.” *IEEE Signal Processing Magazine* 19(1):12–16.

[12] Tibshirani, Robert. 1996. “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58(1):267–88.

[13] Yang, Ling, Ioana Cezara Ene, Reza Arabi Belaghi, David Koff, Nina Stein, and Pasqualina Santaguida. 2022. “Stakeholders’ Perspectives on the Future of Artificial Intelligence in Radiology: A Scoping Review.” *European Radiology* 32(3):1477–95.

[14] Zhu, Yi-Nan, Hong Wu, Chen Cao, and Hai-Ning Li. 2008. “Correlations between Mid-Infrared, Far-Infrared, Hα, and FUV Luminosities for Spitzer SWIRE Field Galaxies.” *The Astrophysical Journal* 686(1):155.