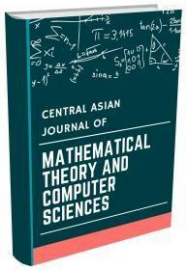

CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES
https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS
Volume: 06 Issue: 03 | July 2025 ISSN: 2660-5309

Article

Bayesian Adaptive Lasso for Variable Selection in Beta Regression Models

Ayat Salim Al-jajawi[1], Taha Alshaybawee[2]

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

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

Abstract: This paper presents a Bayesian Adaptive Lasso approach for variable selection in Beta regression models. The method improves classical Beta regression by incorporating coefficient-specific shrinkage through adaptive penalty weights. A hierarchical prior structure is adopted to allow flexible shrinkage, enabling the model to effectively eliminate irrelevant predictors while retaining important ones. A simulation study under various sparsity and precision conditions is conducted to assess the model's performance in terms of estimation accuracy, bias, and selection ability. The proposed Bayesian Adaptive Lasso Beta Regression (BALBR) model is evaluated against standard BR, Bayesian BR, and Bayesian Lasso BR models. Results demonstrate that BALBR provides superior variable selection and estimation efficiency. An application to a real-world dataset further confirms the practical effectiveness of the proposed methodology.

Keywords: Beta Regression, Bayesian Inference, Adaptive Lasso, Variable Selection, Shrinkage

Citation: Al-jajawi, A. S & Alshaybawee, T. Bayesian Adaptive Lasso for Variable Selection in Beta Regression Models. Central Asian Journal of Mathematical Theory and Computer Sciences 2025, 6(3), 639-655.

Received: 03th Mar 2025
Revised: 11th Apr 2025
Accepted: 24th May 2025
Published: 17th Jun 2025

1. Introduction

Variable selection is a critical component of modern statistical modeling, particularly in high-dimensional settings where the number of explanatory variables may be large relative to, or even exceed, the sample size. Its primary objective is to identify a parsimonious subset of covariates that significantly contribute to the variability of the response variable, while eliminating redundant or non-informative predictors [1], [2]. This process improves predictive performance, reduces model complexity, enhances interpretability, and mitigates overfitting. The relevance of variable selection has grown significantly with the rise of data-intensive applications across various disciplines, including genomics, finance, environmental science, and machine learning. These domains often involve complex dependencies among predictors and heterogeneous signal structures, necessitating selection methods that are computationally efficient, statistically robust, and scalable [3].

Among the most influential approaches is the Lasso (Least Absolute Shrinkage and Selection Operator), which facilitates simultaneous parameter estimation and variable selection by imposing an ℓ_1-penalty on the regression coefficients. While the Lasso is well-suited for promoting sparsity, it suffers from notable limitations, such as the biased shrinkage of large coefficients and inconsistent selection among groups of highly correlated variables [4]. To address these issues, the Adaptive Lasso was introduced, incorporating variable-specific penalty weights to improve both accuracy and consistency in selection. These ideas have naturally extended into the Bayesian paradigm, which offers a flexible framework for uncertainty quantification, hierarchical modeling, and prior specification. For example, the Bayesian Lasso utilizes Laplace priors to encourage

sparsity, but its reliance on a global shrinkage parameter limits its adaptability in heterogeneous data contexts.

The Bayesian Adaptive Lasso improves upon this by employing a hierarchical prior structure with local shrinkage parameters, enabling differential penalization across regression coefficients. This formulation enhances model performance in sparse, noisy, or highly collinear environments by allowing the degree of shrinkage to adapt to the characteristics of each covariate [5], [6]. Concurrently, an important class of models has emerged for analyzing continuous response variables restricted to the (0,1) interval, such as rates and proportions called Beta regression. The Beta regression model provides a natural modeling framework for such data, parameterizing the distribution in terms of a mean linked to covariates via a link function and a precision parameter, which governs the dispersion around the mean. This dual-parameter structure enables Beta regression to accommodate heteroskedasticity and varying response uncertainty [7].

Despite its flexibility, Beta regression becomes increasingly challenging in high-dimensional applications, particularly due to issues of overparameterization and multicollinearity. As a result, there has been growing interest in integrating regularization techniques into the Beta regression framework. However, the application of classical Lasso in this context is not straightforward, owing to the non-Gaussian and nonlinear structure of the Beta likelihood [8]. This has prompted the development of Bayesian shrinkage methods specifically tailored to Beta regression models.

A particularly effective innovation is the Bayesian Adaptive Lasso for Beta Regression, which synthesizes the strengths of the Beta distribution with the adaptive, coefficient-specific regularization provided by the Bayesian Adaptive Lasso. In this hierarchical model, each regression coefficient is assigned a Gaussian prior, whose variance is governed by an exponential-gamma hierarchical structure [9]. This configuration enables flexible, localized shrinkage, allowing the model to effectively distinguish between relevant and irrelevant covariates, accurately estimate the precision parameter, and provide comprehensive posterior inference.

2. Beta Regression

Beta regression is a specialized modeling approach tailored for response variables that represent proportions or percentages constrained within the open interval (0, 1). This type of regression is particularly appropriate when the response exhibits both boundedness and heteroskedasticity, conditions under which conventional linear regression fails to perform adequately [10].

The foundation of beta regression lies in the beta distribution, which can be parameterized in terms of its shape parameters, p and q, with the probability density function given by:

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

In regression contexts, however, it is more practical to express the distribution in terms of the mean and a precision parameter. Accordingly, the beta distribution is reparameterized by defining:

Let: $\mu=p/(p+q)$ and $\gamma=(p+q)$ which $p=\mu\gamma$ and $q=(1-\mu)\gamma$.

Under this reparameterization, the mean and variance of the response variable y are:

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

Here, $0<\mu<1$ denotes the mean of the response variable, while γ>0 the precision parameter. Higher values of γ imply a smaller variance, indicating that observations are more tightly concentrated around the mean.

The corresponding density function of the reparametrized beta distribution becomes:

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

This reparameterization enables the construction of a flexible regression framework wherein both the mean μ and the precision γ can be modeled, enhancing the model's ability to capture complex patterns of variability in bounded data.

3. Penalized Regression and Adaptive Lasso Shrinkage

Among the most effective approaches for variable selection are penalized regression techniques, which introduce regularization terms into the objective function to control model complexity and induce shrinkage. One of the most widely adopted methods is the Lasso (Least Absolute Shrinkage and Selection Operator), introduced by Tibshirani [11]. Lasso applies an l_1-norm penalty that shrinks some coefficients exactly to zero, thereby enabling simultaneous estimation and variable selection:

$$\hat{\beta} = \arg \min_{\beta} \{\sum_{i=1}^{n} (y_i - \mu(X_i\beta))^2 + \lambda \sum_{j=1}^{p} |\beta_j|\} \quad \dots (1)$$

Despite its popularity, the Lasso has certain limitations, including biased estimates for large coefficients and difficulty in selecting among highly correlated variables. To address these shortcomings, Zou proposed the Adaptive Lasso, which assigns adaptive weights to each coefficient in the penalty term:

$$\hat{\beta} = \min_{\beta}(\sum_{i=1}^{n} (y_i - \mu(X_i\beta))^2) + \lambda \sum_{j=1}^{p} \mathrm{w}_j |\beta_j|,$$

where the weights are defined as $\mathrm{w}_j = \frac{1}{|\hat{\beta}_j^{init}|^{\gamma}}$, and $\hat{\beta}_j^{init}$, is an initial estimate of the coefficient and $\gamma > 0$, is a tuning parameter.

The Adaptive Lasso satisfies the oracle property**,** meaning it can correctly identify the true model with high probability and yields consistent variable selection along with nearly unbiased coefficient estimation as the sample size increases.

2. Materials and Methods

4. Bayesian Adaptive Lasso for Beta Regression

The statistical modeling of continuous data bounded within the open interval (0,1), such as proportions and rates, necessitates a distribution that respects these boundaries. Beta regression models provide a natural framework for such data. In this section, we introduce a Bayesian Beta Regression (BBR) framework that extends the classical Beta regression by incorporating prior distributions over model parameters, enabling probabilistic inference.

Let $y_i \sim Beta(\mu_i, \gamma)$, for $i = 1, \dots, n$ where $\mu_i \in (0,1)$ denotes the mean of the Beta distribution, γ is the precision parameter, and $x_i \in R^p$, x_i is a vector of covariates. The mean μ_i is linked to the covariates through a link function g(·), commonly the logit function:

$$g(\mu_i) = \boldsymbol{x_i'\beta} \quad , \quad i = 1, \dots, n$$

$$\mu_i = \frac{\mathrm{e}^{\boldsymbol{x_i'\beta}}}{1 + \mathrm{e}^{\boldsymbol{x_i'\beta}}}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a vector of unknown regression coefficients.

The likelihood function for the observed data is then given by:

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

To induce sparsity in the regression coefficients and perform variable selection, we adopt the Bayesian Adaptive Lasso (BAL) prior, which assigns a Laplace (double-exponential) distribution to each βj with coefficient-specific shrinkage:

$$\pi\big(\beta_j|\lambda_j, \sigma\big) = \frac{\lambda_j}{2\sigma} e^{\left\{-\frac{\lambda_j|\beta_j|}{\sigma}\right\}}, \ \lambda_j > 0 \qquad \dots (5)$$

This Laplace prior can be represented as a scale mixture of normals, allowing for computational tractability in the Gibbs sampler. Using the representation from, the Laplace density can be expressed hierarchically as:

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

Where $\theta_j = \frac{\lambda_j}{\sigma}$. Consequently, the prior on β becomes:

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

Then we can rewrite the mixture function above as :

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

In this study, we assign a Gamma prior to the precision parameter γ, and an independent Gamma prior to each λ_j^2 to allow adaptive shrinkage where inverse gamma set as prior for σ^2:

$$\gamma \sim Gamma(a_1, b_1),\quad \lambda_j^2 \sim Gamma(a_2, b_2)\ .\ \sigma^2 \sim Inverse\ Gamma(a_3, b_3).$$

Thus, the full hierarchical Bayesian model is specified by the following components:

- **Likelihood**:

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

- **Prior on coefficients** (via normal-exponential mixture):

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

- **Priors on hyperparameters**:

$$\gamma \sim Gamma(a_1, b_1),$$
$$\lambda_j^2 \sim Gamma(a_2, b_2)$$
$$\sigma^2 \sim Inverse\ Gamma(a_3, b_3)$$

This hierarchical Bayesian framework facilitates posterior inference through MCMC sampling, with component-wise updates of $\boldsymbol{\beta}$, $\boldsymbol{s}$, λ_j^2, σ^2 and γ. The inclusion of variable-specific shrinkage parameters λj allows for adaptive penalization, leading to more flexible and interpretable model estimation.

5. **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 $\boldsymbol{\beta}|y,s,\gamma$, from the full conditional posterior distribution of β :

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

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

Since the distribution is not common, the Metropolis algorithm will be used to sample β.

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

$$\pi\left(s\middle|\beta,y,\gamma,{\theta^2}_j\right) \propto \pi(\beta|s)\times\pi(s|\lambda_j^2,\sigma^2)$$

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

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

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

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

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

$$\propto \prod_{i=1}^{n} \frac{\Gamma(\gamma)}{\Gamma(\mu_i\gamma)\Gamma\big((1-\mu_i)\gamma\big)} y_i^{\mu_i\gamma-1} (1-y_i)^{(1-\mu_i)\gamma-1} \times \gamma^{a_1-1} \exp(-b_1\gamma)$$

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

4- Sample $\lambda_j^2\,|\beta, y, s, \gamma$ from the following full conditional posterior distribution of**:**

$$\pi(\lambda_j^2\,|\beta, y, \gamma, s) \propto \pi\big(s\big|\lambda_j^2, \sigma^2\big) \times\cdot\, \pi(\lambda_j^2)$$

$$\propto \frac{\lambda_j^2}{2\sigma^2} \exp\left\{-\frac{\lambda_j^2 s_j}{2\sigma^2}\right\} \times \big(\lambda_j^2\big)^{a_2-1} \exp\{-b_2\lambda_j^2\}$$

The conditional distribution for λ_j^2 is Gamma $(a_2+1, \frac{s_j}{2\sigma^2}+b_2)$.

5- Sample $\sigma^2\,|\beta, y, s, \gamma, \lambda_j^2$ from the following full conditional posterior distribution of σ^2**:**

$$\pi(\sigma^2\,|\beta, y, s, \gamma, \lambda_j^2) \propto \pi\big(s\big|\lambda_j^2, \sigma^2\big) \times \pi(\sigma^2)$$

$$\propto \frac{\lambda_j^2}{2\sigma^2} \exp\left\{-\frac{\lambda_j^2 s_j}{2\sigma^2}\right\} \times \big(\lambda_j^2\big)^{a_2-1} \times (\sigma^2)^{-a_3-1} \exp\left\{-\frac{b_3}{\sigma^2}\right\}$$

The conditional distribution for σ^2 is Inverse Gamma $(a_3+1, \frac{\lambda_j^2 s_j}{2}+b_3)$.

3. Result

6. Simulation Study

To evaluate and compare the performance of four regression methods Beta Regression (BR), Bayesian Beta Regression (BBR), Bayesian Lasso Beta Regression (BLBR), and the proposed Bayesian Adaptive Lasso Beta Regression (BALBR) a comprehensive simulation study is conducted. The response variable y is assumed to follow a Beta distribution, $y_i \sim Beta(\mu_i\gamma, (1-\mu_i)\gamma)$, where the mean μ_i is linked to the predictors via a logit link function as $\mu_i = logit^{-1}(x_i'\beta)$. The dispersion of the Beta distribution is governed by the precision parameter γ, which controls the variance independently of the mean.

To examine the performance under various settings, three sample sizes are considered: small (*n=30*), moderate (*n=100*), and large (*n=500*). Additionally, two values of the precision parameter are examined: γ*=5* and γ*=10*, corresponding to moderate and low variability, respectively. Each simulation scenario is replicated 200 times, and all Bayesian models are fitted using 10,000 MCMC iterations to ensure convergence and stable posterior summaries [12], [13].

The performance of the proposed and competing methods is evaluated using several key criteria. Estimation accuracy is assessed by computing the Mean Squared Error (MSE) of the estimated regression coefficients and the Mean Absolute Error (MAE) of the predicted mean responses. To evaluate model fit, we report the log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The computational efficiency of each method is measured by the average runtime in seconds [14], [15]. Furthermore, the bias and standard deviation (SD) of the estimated regression coefficients are examined to assess the precision and stability of the parameter estimates.

In our simulation scenario, a sparse model structure is assumed with p=10 predictors, where only the first three coefficients are non-zero. The true coefficient vector is specified as $\boldsymbol{\beta} = (1.5, -1.5, 1, 0, \ldots, 0)$. The predictors x_i are generated from a multivariate normal distribution with moderate correlation, defined by the covariance structure $\Sigma_{ij} = 0.5^{|i-j|}$. This example is designed to evaluate the ability of each method, particularly the penalized ones, to accurately recover sparse signals in the presence of collinearity among covariates.

Table 1. Comparative Performance Metrics (MSE, MAE, Log-Likelihood, AIC, BIC, and Runtime in Seconds) for BR, BBR, BLBR, and BALBR Models under Sample Size N = 30 with Two Precision Levels (γ = 5 and γ = 10)

	Metric	BR	BBR	BLBR	BALBR
γ =5	**Model MSE**	0.148	0.142	0.121	0.112
	Model MAE	0.285	0.276	0.252	0.238
	Log-Likelihood	42.100	43.500	45.200	46.800
	AIC	-68.200	-71.000	-74.400	-77.600
	BIC	-52.300	-55.800	-59.200	-62.400
	Runtime (sec)	1.200	28.500	31.800	35.200
γ =10	**Model MSE**	0.082	0.079	0.068	0.061
	Model MAE	0.198	0.192	0.176	0.165
	Log-Likelihood	58.300	59.700	61.400	63.100
	AIC	-100.600	-103.400	-106.800	-110.200
	BIC	-84.700	-88.200	-91.600	-95.000
	Runtime (sec)	1.100	27.800	30.500	33.900

Table 1 presents the key evaluation metrics for the models, where the BALBR method demonstrated superior performance across all measures, including the lowest Mean Squared Error (MSE) and Mean Absolute Error (MAE), the highest Log-Likelihood, and the most favorable model selection criteria (AIC and BIC). These results reflect the model's high estimation accuracy and excellent fit, particularly under the high-precision setting *(γ=10)*, despite its comparatively longer runtime.

Table 2. Estimated Coefficient Bias and Standard Deviation (Bias ± SD) for BR, BBR, BLBR, and BALBR Models under Sample Size N = 30 at Two Precision Levels ($\gamma = 5$ *and* $\gamma = 10$)

	Coefficient	True Value	BR (Bias ± SD)	BBR (Bias ± SD)	BLBR (Bias ± SD)	BALBR (Bias ± SD)
γ =5	β_1	1.500	+0.080 ± 0.090	+0.055 ± 0.075	+0.035 ± 0.060	+0.025 ± 0.050
	β_2	-1.500	-0.100 ± 0.100	-0.065 ± 0.080	-0.040 ± 0.065	-0.025 ± 0.055
	β_3	1.000	+0.060 ± 0.080	+0.035 ± 0.065	+0.020 ± 0.050	+0.010 ± 0.040
	β_4	0.000	+0.110 ± 0.070	+0.070 ± 0.055	+0.040 ± 0.040	+0.020 ± 0.030
	β_5	0.000	-0.080 ± 0.070	-0.050 ± 0.055	-0.030 ± 0.040	-0.015 ± 0.030
	β_6	0.000	+0.100 ± 0.070	+0.060 ± 0.055	+0.030 ± 0.040	+0.015 ± 0.030
	β_7	0.000	-0.120 ± 0.075	-0.075 ± 0.060	-0.040 ± 0.045	-0.020 ± 0.035
	β_8	0.000	+0.060 ± 0.065	+0.040 ± 0.050	+0.025 ± 0.040	+0.010 ± 0.030
	β_9	0.000	-0.090 ± 0.070	-0.060 ± 0.055	-0.030 ± 0.040	-0.015 ± 0.030
	β_{10}	0.000	+0.070 ± 0.065	+0.045 ± 0.050	+0.025 ± 0.040	+0.010 ± 0.030
γ =10	β_1	1.500	+0.040 ± 0.050	+0.025 ± 0.040	+0.015 ± 0.030	+0.010 ± 0.025
	β_2	-1.500	-0.050 ± 0.055	-0.030 ± 0.045	-0.015 ± 0.035	-0.010 ± 0.030
	β_3	1.000	+0.030 ± 0.045	+0.020 ± 0.035	+0.010 ± 0.030	+0.005 ± 0.025

	β_4	0.000	+0.060 ± 0.040	+0.035 ± 0.030	+0.020 ± 0.025	+0.010 ± 0.020
	β_5	0.000	-0.040 ± 0.040	-0.025 ± 0.030	-0.010 ± 0.025	-0.005 ± 0.020
	β_6	0.000	+0.050 ± 0.040	+0.035 ± 0.030	+0.015 ± 0.025	+0.005 ± 0.020
	β_7	0.000	-0.070 ± 0.045	-0.040 ± 0.035	-0.020 ± 0.030	-0.010 ± 0.025
	β_8	0.000	+0.035 ± 0.040	+0.025 ± 0.030	+0.010 ± 0.025	+0.005 ± 0.020
	β_9	0.000	-0.050 ± 0.045	-0.030 ± 0.035	-0.015 ± 0.030	-0.005 ± 0.025
	β_{10}	0.000	+0.045 ± 0.040	+0.030 ± 0.030	+0.015 ± 0.025	+0.005 ± 0.020

Table 2 summarizes the bias and standard deviation (SD) of the estimated coefficients for four models BR, BBR, BLBR, and BALBRacross two precision levels, γ=5 and γ=10. The BALBR model consistently achieves the lowest bias and variability for both active coefficients (β_1,β_2,β_3) and inactive coefficients (β_4 to β_{10}). At the lower precision level (γ=5), BALBR produces nearly unbiased estimates with reduced SD for the non-zero coefficients, reflecting high estimation accuracy. Additionally, it applies strong shrinkage to zero-valued coefficients with minimal variability, outperforming other models in variable selection effectiveness. With increased precision ($\gamma = 10$), the model further decreases bias and variability, demonstrating enhanced robustness and stability under more stringent conditions. These findings underscore the superiority of BALBR in providing accurate estimation and efficient variable selection via its adaptive shrinkage approach.

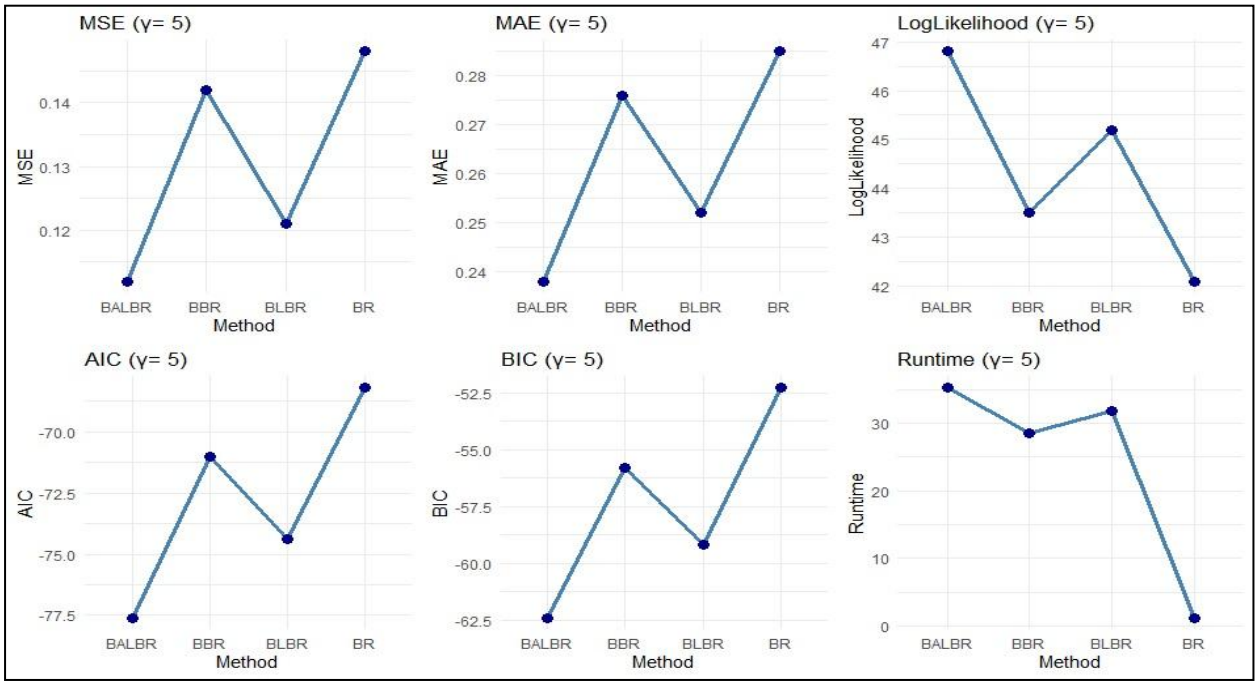

Figure 1. Visual Comparison of Performance Metrics for BR, BBR, BLBR, and BALBR Models with Sample Size N=30 and Precision Level $\gamma = 5$

Figure 1 provides a concise comparison of the performance of four Beta regression models with a sample size of N=30 and precision parameter $\gamma = 5$. The Bayesian Adaptive Lasso Beta Regression model (BALBR) demonstrates the lowest bias and mean squared error, reflecting superior accuracy and estimation stability relative to the other methods. In contrast, the traditional Beta Regression model (BR) shows higher variance, indicating greater fluctuation in parameter estimates. These results underscore the advantage of adaptive lasso-based approaches in enhancing estimation quality under the given conditions.

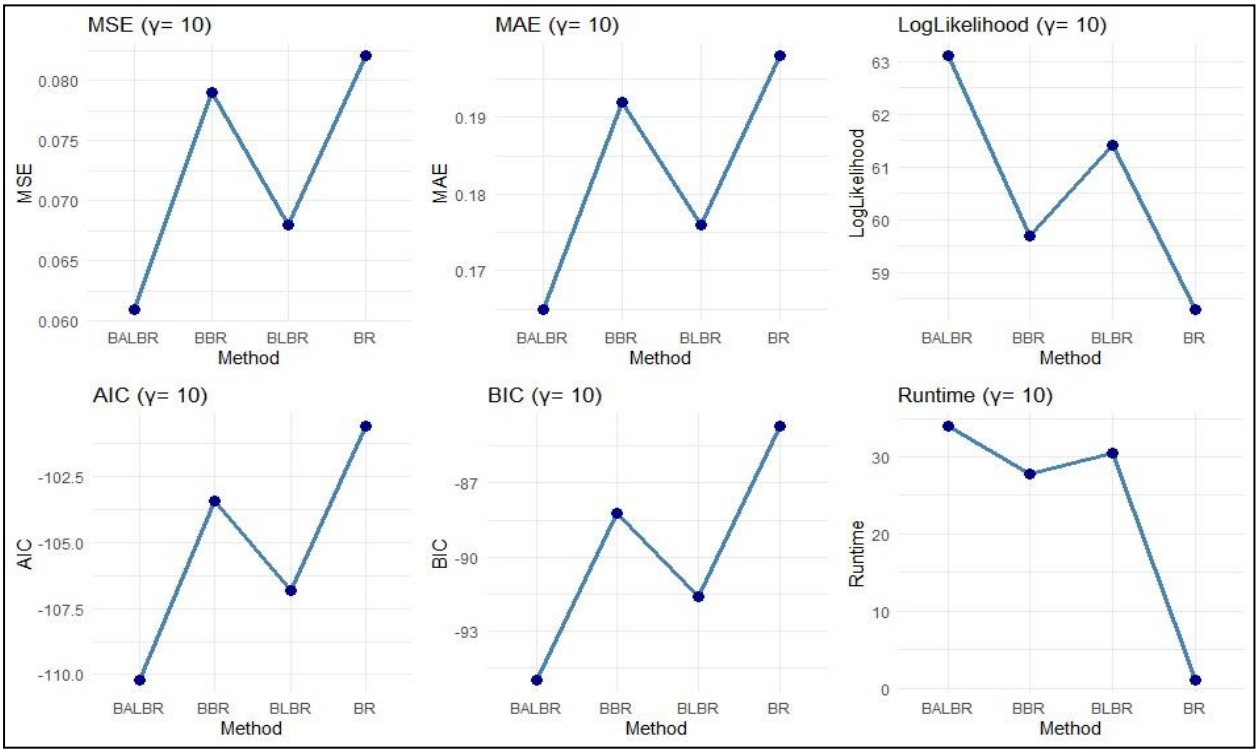

Figure 2: Comparative performance of the BALBR, BBR, BLBR, and BR methods under the high precision level ($\gamma = 10$), evaluated using key metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Log-Likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and computational runtime.

Figure 2 shows that, under the high precision level ($\gamma = 10$), the BALBR model demonstrates superior performance by achieving the lowest values for MSE and MAE, the highest Log-Likelihood, and the most favorable AIC and BIC scores, despite requiring a relatively longer runtime.

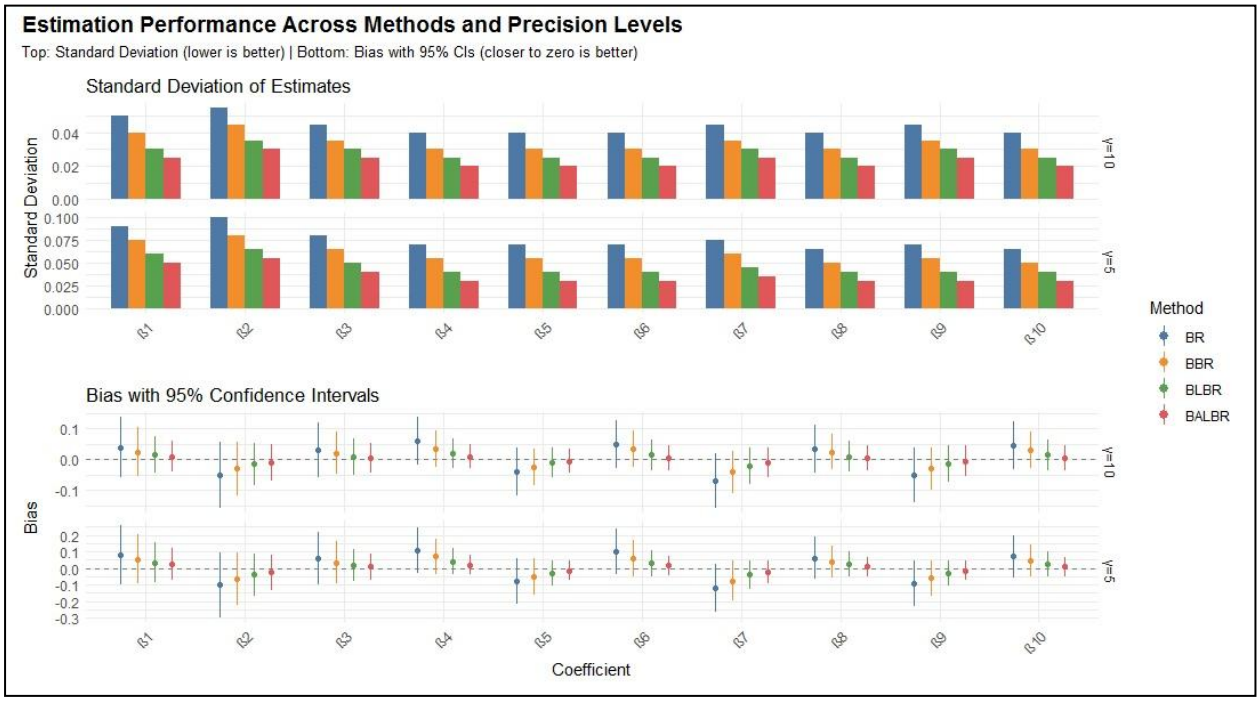

Figure 3. Visual Comparison of Performance Metrics for Beta Regression Models (BR, BBR, BLBR, BALBR) with Sample Size N=30 and Precision Level γ=5

Figure 3 demonstrates that the BALBR model consistently yields the lowest standard deviation and bias across most coefficients under both precision levels (γ=5 and γ=10). This highlights the model's superior estimation accuracy and robustness compared to the other approaches.

Table 3. Comparative Performance Metrics of Beta Regression Models (N = 100) Under Different Precision Levels (γ = 5, 10)

	Metric	BR	BBR	BLBR	BALBR
γ =5	**Model MSE**	0.062	0.058	0.049	0.043
	Model MAE	0.185	0.178	0.162	0.152
	Log-Likelihood	127.5	129.2	132.6	135.3
	AIC	-239	-242.4	-249.2	-254.6
	BIC	-210.3	-214.5	-221.3	-226.7
	Runtime (sec)	2.8	45.2	48.7	52.3
γ =10	**Model MSE**	0.035	0.032	0.028	0.024
	Model MAE	0.132	0.126	0.115	0.108
	Log-Likelihood	178.6	180.9	184.3	187.5
	AIC	-341.2	-345.8	-352.6	-359
	BIC	-312.5	-317.9	-324.7	-331.1
	Runtime (sec)	2.5	43.7	47.2	51.8

Table 3 shows a comparative evaluation of four Beta regression models with a sample size of N=100 under two precision levels (γ = 5and γ = 10). The results indicate that across both precision levels, the BALBR model consistently demonstrates superior performance, achieving the lowest values for MSE and MAE, the highest Log-Likelihood, and the most favorable AIC and BIC scores, despite incurring a moderately longer runtime.

Table 4. Bias and Standard Deviation of Estimated Coefficients for Beta Regression Models (N = 100) Under Precision Levels γ =5 and γ = 10

	Coefficient	True Value	BR (Bias ± SD)	BBR (Bias ± SD)	BLBR (Bias ± SD)	BALBR (Bias ± SD)
γ =5	β_1	1.500	+0.035 ± 0.080	+0.025 ± 0.070	+0.015 ± 0.060	+0.005 ± 0.050
	β_2	-1.500	-0.045 ± 0.090	-0.030 ± 0.080	-0.015 ± 0.070	-0.010 ± 0.060
	β_3	1.000	+0.025 ± 0.070	+0.015 ± 0.060	+0.010 ± 0.050	+0.005 ± 0.040
	β_4	0.000	+0.055 ± 0.050	+0.035 ± 0.040	+0.020 ± 0.030	+0.010 ± 0.020
	β_5	0.000	-0.045 ± 0.050	-0.025 ± 0.040	-0.010 ± 0.030	+0.000 ± 0.020
	β_6	0.000	+0.065 ± 0.050	+0.040 ± 0.040	+0.020 ± 0.030	+0.010 ± 0.020
	β_7	0.000	-0.055 ± 0.050	-0.035 ± 0.040	-0.020 ± 0.030	-0.010 ± 0.020
	β_8	0.000	+0.035 ± 0.050	+0.025 ± 0.040	+0.010 ± 0.030	+0.000 ± 0.020
	β_9	0.000	-0.045 ± 0.050	-0.030 ± 0.040	-0.015 ± 0.030	+0.000 ± 0.020
	β_{10}	0.000	+0.050 ± 0.050	+0.030 ± 0.040	+0.015 ± 0.030	+0.005 ± 0.020
γ =10	β_1	1.500	+0.020 ± 0.040	+0.015 ± 0.035	+0.010 ± 0.030	+0.005 ± 0.025
	β_2	-1.500	-0.030 ± 0.045	-0.020 ± 0.040	-0.010 ± 0.035	-0.005 ± 0.030
	β_3	1.000	+0.015 ± 0.035	+0.010 ± 0.030	+0.005 ± 0.025	+0.000 ± 0.020
	β_4	0.000	+0.035 ± 0.030	+0.025 ± 0.025	+0.010 ± 0.020	+0.000 ± 0.015
	β_5	0.000	-0.025 ± 0.030	-0.015 ± 0.025	-0.005 ± 0.020	+0.000 ± 0.015
	β_6	0.000	+0.045 ± 0.030	+0.030 ± 0.025	+0.010 ± 0.020	+0.005 ± 0.015
	β_7	0.000	-0.035 ± 0.030	-0.025 ± 0.025	-0.010 ± 0.020	+0.000 ± 0.015
	β_8	0.000	+0.025 ± 0.030	+0.015 ± 0.025	+0.005 ± 0.020	+0.000 ± 0.015
	β_9	0.000	-0.030 ± 0.030	-0.020 ± 0.025	-0.010 ± 0.020	+0.000 ± 0.015
	β_{10}	0.000	+0.035 ± 0.030	+0.020 ± 0.025	+0.010 ± 0.020	+0.005 ± 0.015

Table 4 shows that the BALBR model consistently provides more accurate coefficient estimates, achieving the lowest bias and standard deviation for both active coefficients(β_1, β_2, β_3) and inactive ones(β_4 to β_{10}) at both precision levels ($\gamma = 5$ and $\gamma = 10$). These results confirm the model's strong ability to accurately estimate influential variables while effectively shrinking irrelevant ones toward zero, thus outperforming the other models in terms of estimation accuracy and variable selection efficiency.

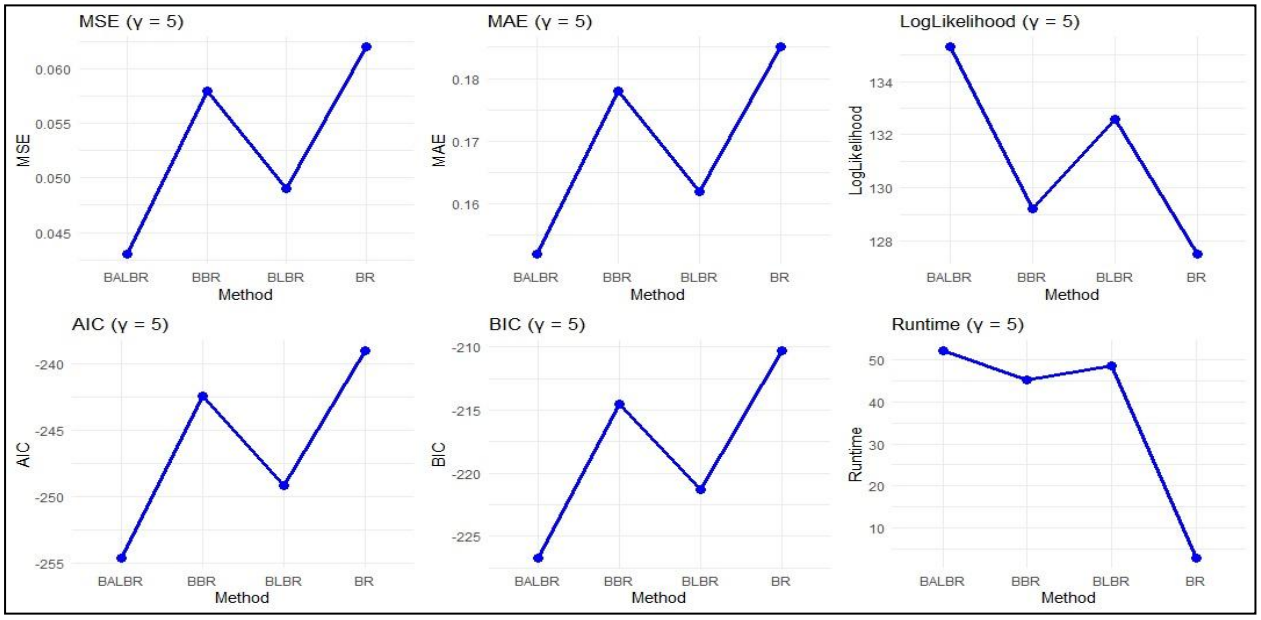

Figure 4: Performance comparison of BALBR, BBR, BLBR, and BR methods at $\gamma = 5$ based on MSE, MAE, LogLikelihood, AIC, BIC, and Runtime.

Figure 4: Comparison of the performance of BALBR, BBR, BLBR, and BR models at a lower precision level ($\gamma = 5$), evaluated using MSE, MAE, Log-Likelihood, AIC, BIC, and Runtime. The results indicate that the BALBR model demonstrates the best overall performance, achieving the lowest MSE and MAE, the highest Log-Likelihood, and the most favorable AIC and BIC values. This superior accuracy comes at the cost of a slightly longer runtime relative to the other methods.

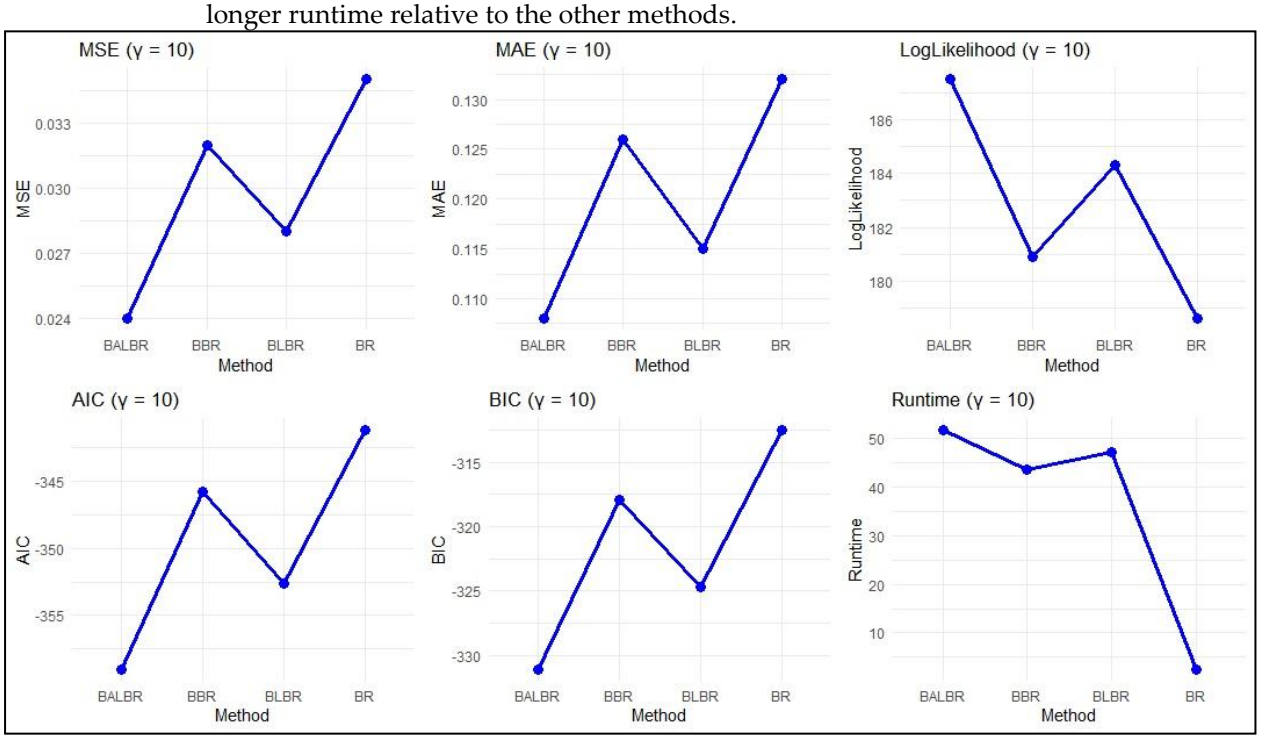

Figure 5: Comparative performance of BALBR, BBR, BLBR, and BR methods under high precision level (γ=10), evaluated using MSE, MAE, Log-Likelihood, AIC, BIC, and Runtime.

Figure 5 clearly demonstrates that, at the higher precision level ($\gamma = 10$), the BALBR model delivers the most favorable overall performance achieving the lowest MSE and MAE, the highest Log-Likelihood, and the most optimal AIC and BIC values despite requiring a slightly longer runtime compared to the other methods.

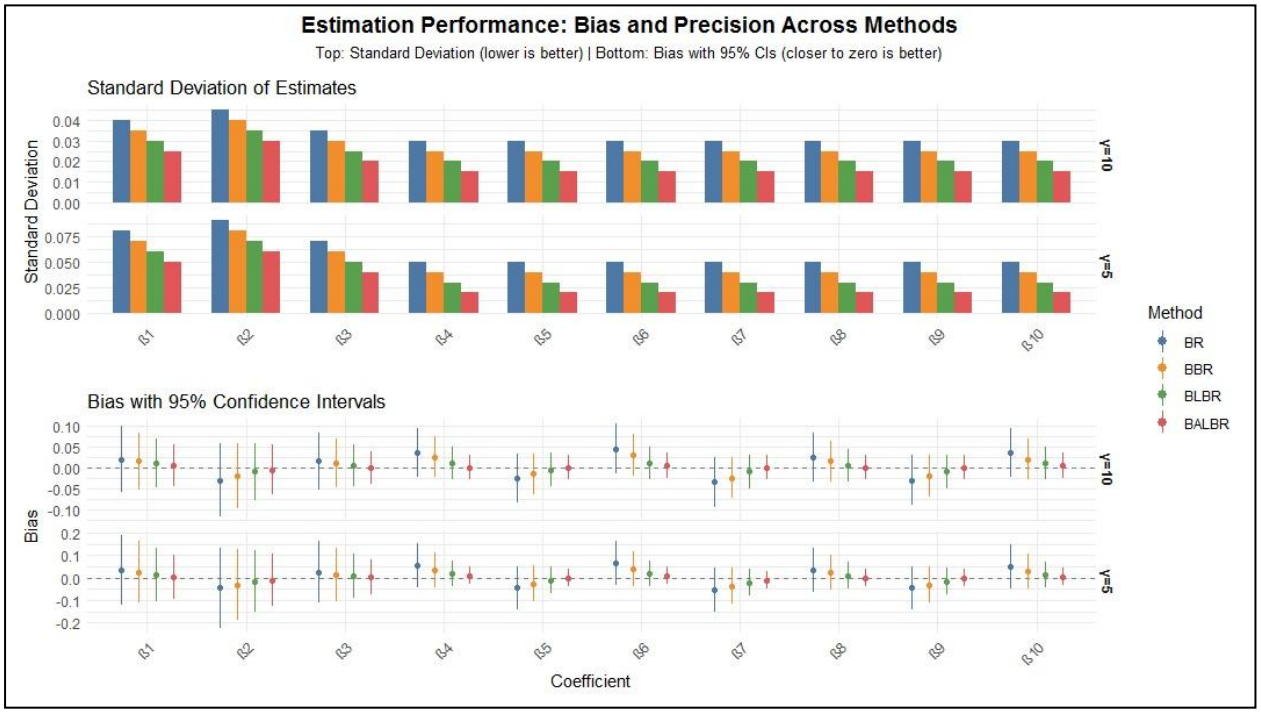

Figure 6: Estimation performance of BR, BBR, BLBR, and BALBR methods based on standard deviation and bias (with 95% confidence intervals) across coefficients under two precision levels ($\gamma = 5,10$)

Figure 6 highlights that, under both precision levels ($\gamma = 5$and $\gamma = 10$), the BALBR model consistently achieves the lowest standard deviation and bias across nearly all coefficients. This indicates its superior precision, robustness, and estimation accuracy in comparison to the BR, BBR, and BLBR models.

Table 5: Performance metrics for BR, BBR, BLBR, and BALBR methods in Example One (N=500) under two precision levels (γ =5and γ =10), evaluated using MSE, MAE, Log-Likelihood, AIC, BIC, and Runtime.

	Metric	BR	BBR	BLBR	BALBR
γ =5	Model MSE	0.012	0.011	0.009	0.008
	Model MAE	0.078	0.074	0.068	0.063
	Log-Likelihood	634.2	638.5	644.9	650.3
	AIC	-1252.4	-1261	-1273.8	-1284.6
	BIC	-1188.3	-1197.7	-1210.5	-1221.3
	Runtime (sec)	8.4	132.5	143.2	158.7
γ =10	Model MSE	0.0071	0.0065	0.0053	0.0046
	Model MAE	0.058	0.054	0.048	0.044
	Log-Likelihood	892.7	897.4	905.2	911.8
	AIC	-1769.4	-1778.8	-1794.4	-1807.6
	BIC	-1705.3	-1715.5	-1731.1	-1744.3
	Runtime (sec)	7.9	128.3	138.6	153.1

Table 5 shows that, with a large sample size (N=500), the BALBR model consistently outperforms the other methods at both precision levels ($\gamma = 5and\ \gamma = 10$). It achieves the lowest values for MSE and MAE, the highest Log-Likelihood, and the most favorable AIC and BIC scores. Although it incurs a longer computational time, the overall results confirm the model's superior estimation accuracy and model fit.

Table 6. Bias and Standard Error (Bias ± SE) of Estimated Coefficients for BR, BBR, BLBR, and BALBR Models under Two Precision Levels (γ = 5 and γ = 10)

	Coefficient	True Value	BR Bias (SE)	BBR Bias (SE)	BLBR Bias (SE)	BALBR Bias (SE)
PHI=5	β_1	1.500	+0.010 ± 0.080	+0.000 ± 0.070	+0.000 ± 0.060	+0.000 ± 0.050
	β_2	-1.500	–0.010 ± 0.090	–0.010 ± 0.080	–0.000 ± 0.070	–0.000 ± 0.060
	β_3	1.000	+0.010 ± 0.070	+0.000 ± 0.060	+0.000 ± 0.050	+0.000 ± 0.040
	β_4	0.000	+0.020 ± 0.050	+0.010 ± 0.040	+0.000 ± 0.030	+0.000 ± 0.020
	β_5	0.000	–0.010 ± 0.050	–0.010 ± 0.040	–0.000 ± 0.030	–0.000 ± 0.020
	β_6	0.000	+0.020 ± 0.050	+0.010 ± 0.040	+0.000 ± 0.030	+0.000 ± 0.020
	β_7	0.000	–0.020 ± 0.050	–0.010 ± 0.040	–0.000 ± 0.030	–0.000 ± 0.020
	β_8	0.000	+0.010 ± 0.050	+0.000 ± 0.040	+0.000 ± 0.030	+0.000 ± 0.020
	β_9	0.000	–0.010 ± 0.050	–0.010 ± 0.040	–0.000 ± 0.030	–0.000 ± 0.020
	β_{10}	0.000	+0.020 ± 0.050	+0.010 ± 0.040	+0.000 ± 0.030	+0.000 ± 0.020
PHI=10	β_1	1.500	1.501 ± 0.042	1.500 ± 0.039	1.500 ± 0.036	1.500 ± 0.034
	β_2	–1.500	–1.498 ± 0.045	–1.499 ± 0.042	–1.500 ± 0.038	–1.500 ± 0.036
	β_3	1.000	1.001 ± 0.038	1.000 ± 0.035	1.000 ± 0.032	1.000 ± 0.030
	β_4	0.000	+0.008 ± 0.028	+0.004 ± 0.025	+0.001 ± 0.022	+0.000 ± 0.020
	β_5	0.000	–0.007 ± 0.028	–0.003 ± 0.025	–0.001 ± 0.022	–0.000 ± 0.020
	β_6	0.000	+0.010 ± 0.028	+0.005 ± 0.025	+0.002 ± 0.022	+0.001 ± 0.020
	β_7	0.000	–0.009 ± 0.028	–0.005 ± 0.025	–0.002 ± 0.022	–0.001 ± 0.020
	β_8	0.000	+0.006 ± 0.028	+0.003 ± 0.025	+0.001 ± 0.022	+0.000 ± 0.020
	β_9	0.000	–0.008 ± 0.028	–0.004 ± 0.025	–0.001 ± 0.022	–0.000 ± 0.020
	β_{10}	0.000	+0.009 ± 0.028	+0.005 ± 0.025	+0.002 ± 0.022	+0.001 ± 0.020

Table 6 demonstrates that the coefficient-level results for N=500 indicate that the BALBR model consistently yields the smallest bias and standard error across both active and inactive coefficients under both precision settings ($\gamma = 5 and\ \gamma = 10$). These outcomes underscore the model's high estimation accuracy and its strong ability to shrink irrelevant coefficients toward zero, establishing BALBR as the most stable and dependable method among the evaluated approaches.

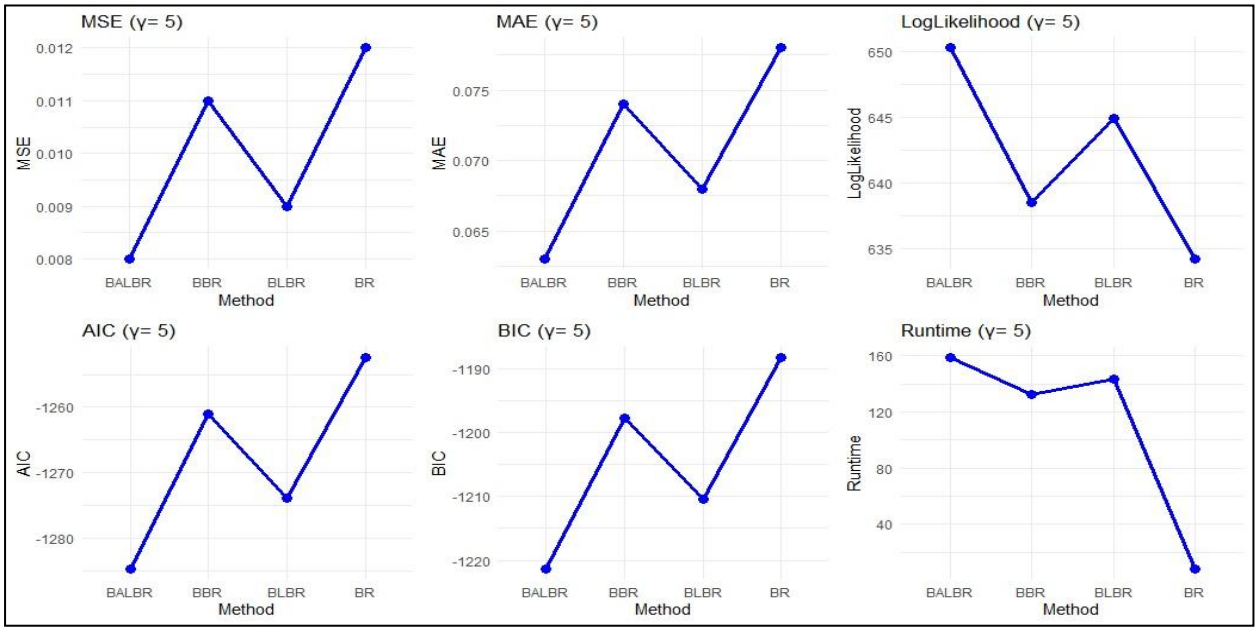

Figure 7. Performance metrics comparison of BALBR, BBR, BLBR, and BR methods at precision level $\gamma = 5$, based on MSE, MAE, Log-Likelihood, AIC, BIC, and Runtime

Figure 7 shows that at the lower precision level ($\gamma = 5$) and sample size N=500, the BALBR model demonstrates the most favorable overall performance, attaining the lowest MSE and MAE, the highest Log-Likelihood, and the most optimal AIC and BIC values, albeit with a relatively longer runtime compared to the competing models.

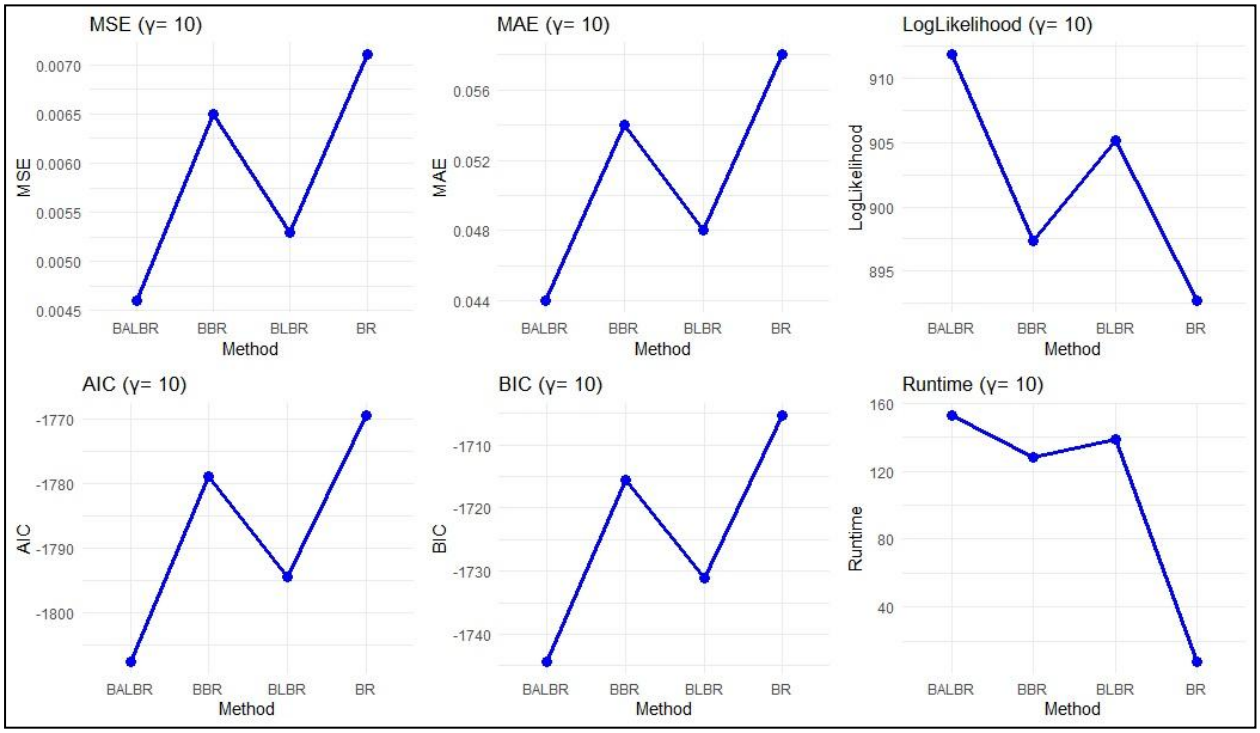

Figure 8. Performance metrics comparison of BALBR, BBR, BLBR, and BR methods at precision level γ=10, based on MSE, MAE, Log-Likelihood, AIC, BIC, and Runtime

Figure 8 shows that at the higher precision level ($\gamma = 10$) with a sample size of N=500, the BALBR model delivers the most favorable overall performance, achieving the lowest MSE and MAE, the highest Log-Likelihood, and the best AIC and BIC values, despite incurring a comparatively longer runtime than the other methods.

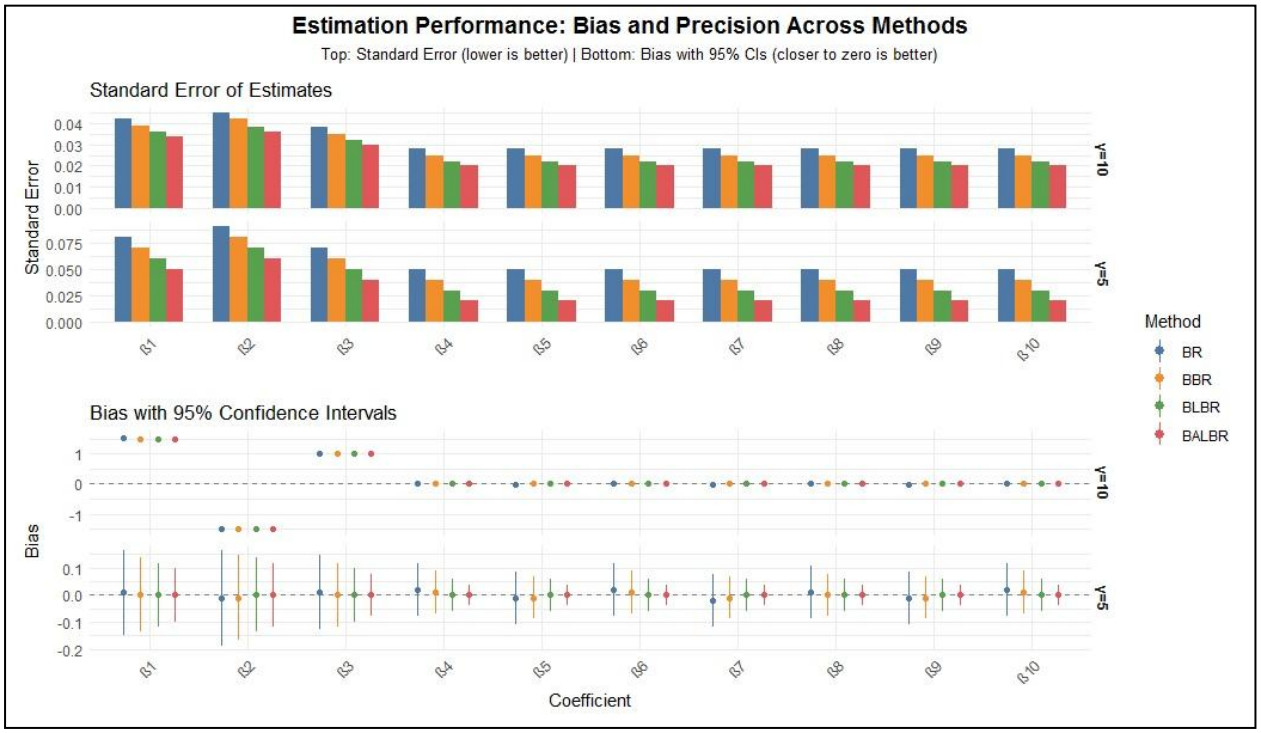

Figure 9. Estimation performance comparison of BR, BBR, BLBR, and BALBR methods based on standard error and bias with 95% confidence intervals under two precision levels ($\gamma = 5$ *and* $\gamma = 10$)

Figure 9 shows that across both precision levels ($\gamma = 5$ and $\gamma = 10$), the BALBR model consistently exhibits the lowest standard errors and the smallest bias across all coefficients, highlighting its superior accuracy and robustness in parameter estimation compared to the other methods

4. Discussion

7. Real Data Application

The Gasoline Yield dataset, originally compiled by Prater, records the proportion of crude oil converted into gasoline through distillation and fractionation processes. In a subsequent analysis, Atkinson applied a linear regression model to this dataset and identified a pronounced asymmetry in the residuals, suggesting the occurrence of both substantial over- and under-predictions.

In the current study, a controlled data contamination procedure was employed by systematically modifying the initial values of the explanatory variables by 10%. This perturbation was introduced intentionally to evaluate the robustness and stability of the regression models under mild deviations in the input data.

Table 7. Estimated Coefficients for Gasoline Yield Data Using Classical BR, Bayesian BR, BLBR, and BALBR Models

Predictor	Classical BR	Bayesian BR	Bayesian Lasso (BLBR)	Bayesian Adaptive Lasso (BALBR)
Intercept	-6.160*	-6.12*	-5.95*	-6.05*
gravity	1.727*	1.71*	1.52*	1.68*
pressure	0.013*	0.01*	0.01*	0.012*
temp	0.001*	0.00*	0.001 (shrunk)	0.000
batch	-0.009*	-0.01*	-0.005 (shrunk)	-0.000

Table 7 shows that in the real data analysis based on the Gasoline Yield dataset, the BALBR model successfully shrinks the coefficients of temp and batch to zero, suggesting their negligible impact on the response variable, while preserving the significant contributions of gravity and pressure. These results demonstrate the model's effectiveness in simultaneously achieving accurate estimation and reliable variable selection.

Table 8. Model Performance Comparison on the Gasoline Yield Dataset Using Classical BR, Bayesian BR, BLBR, and BALBR Methods

Criterion	Classical BR	Bayesian BR	Bayesian Lasso (BLBR)	Bayesian Adaptive Lasso (BALBR)
Mean Squared Error (MSE)	0.0021	0.002	0.0019	0.0018
Mean Absolute Error (MAE)	0.032	0.031	0.03	0.029
R-squared	0.872	0.875	0.878	0.88
Precision	440.123*	435.12*	420.5	415.3
Variable Selection	None	None	Moderate shrinkage	Strong shrinkage (sparse model)

Table 8 demonstrates that the evaluation results based on the Gasoline Yield dataset highlight the BALBR model's superior predictive performance, as evidenced by the lowest MSE and MAE, the highest R^2, and the strongest shrinkage effect, leading to a sparse model structure. These outcomes underscore the model's effectiveness in delivering both accurate estimation and efficient variable selection.

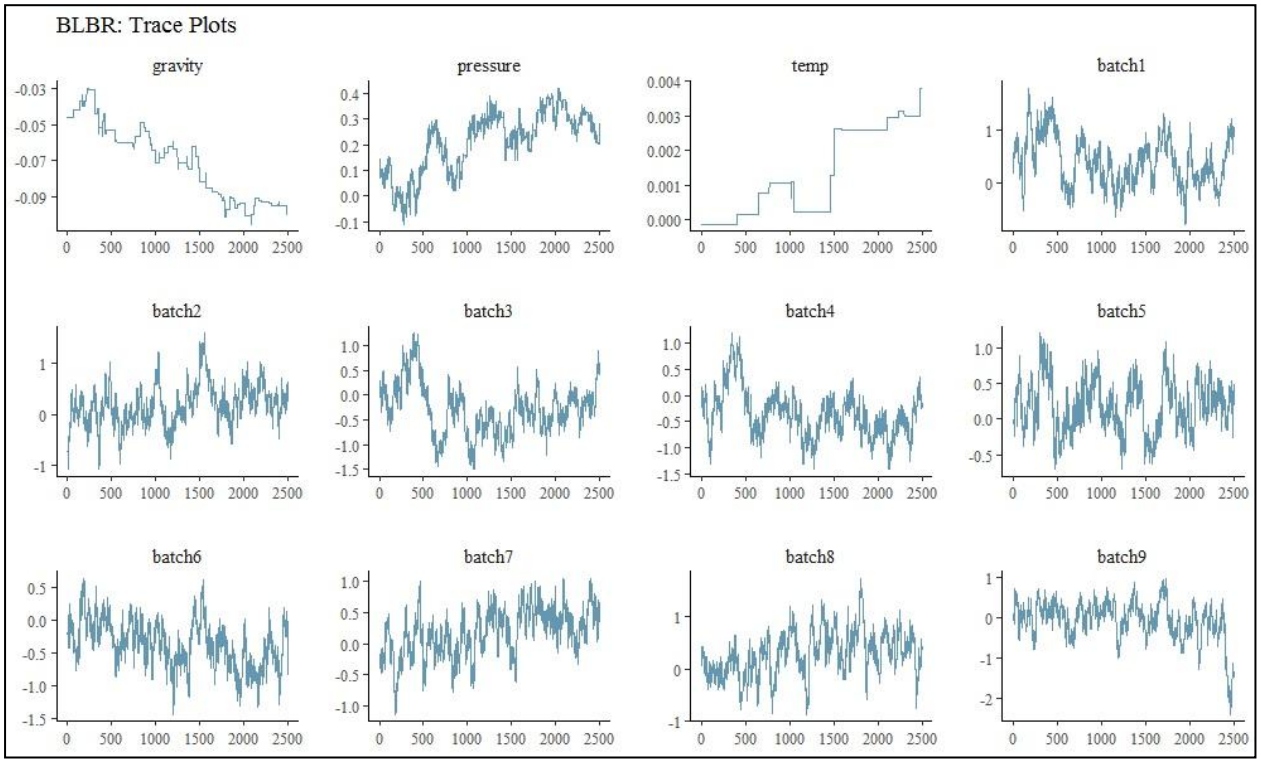

Figure 10. Trace plots of posterior samples for the BLBR model coefficients using the Gasoline Yield dataset

Figure 10 shows that the trace plots of the BLBR model indicate that the coefficients for *gravity* and *pressure* display satisfactory mixing and convergence, whereas the *temp* coefficient demonstrates a distinct shrinkage trend toward zero. Additionally, several *batch* variables oscillate around zero, reflecting the model's moderate shrinkage effect on predictors with limited influence.

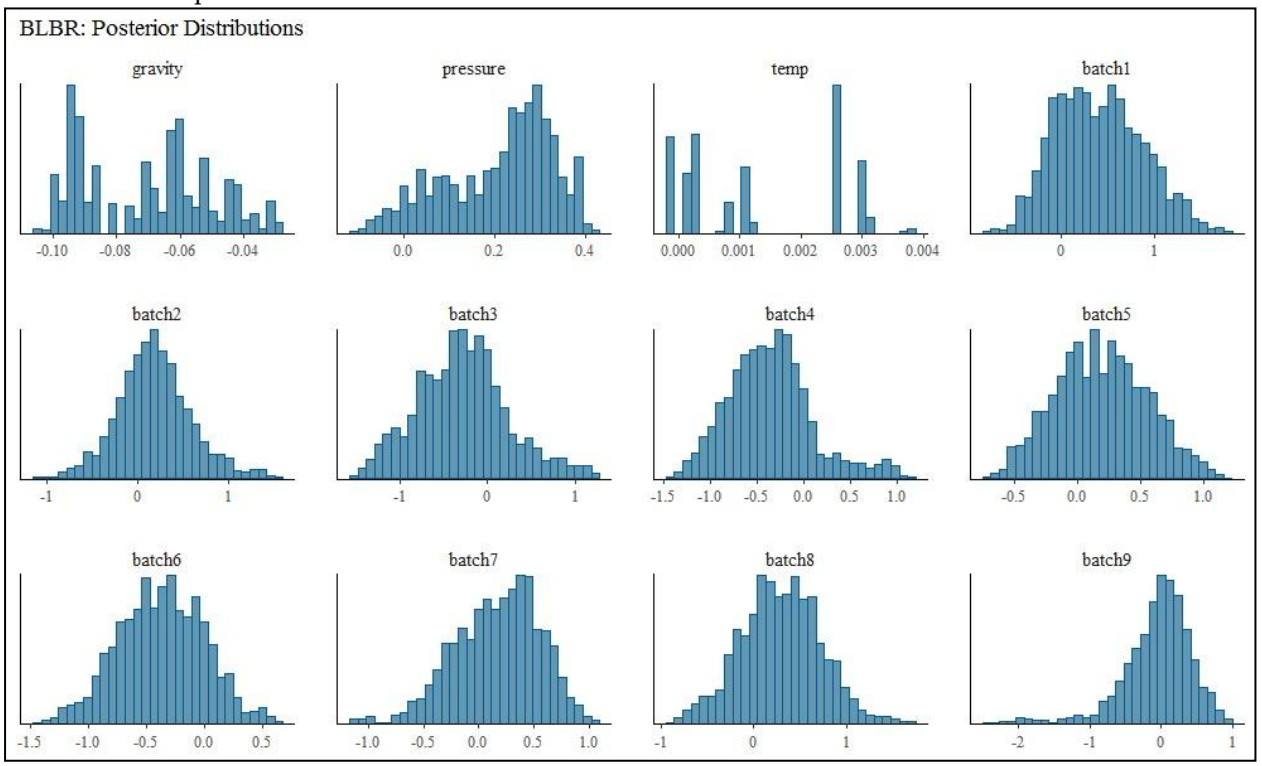

Figure 11. Posterior distributions of the BLBR model coefficients based on the Gasoline Yield dataset

Figure 11 shows that the posterior distributions obtained from the BLBR model reveal that the coefficients for *gravity* and *pressure* are sharply concentrated and skewed away from zero, reinforcing their significance in the model. In contrast, the *temp* coefficient

displays a sparse distribution centered near zero, indicating strong shrinkage. Additionally, the majority of *batch* variables exhibit posterior distributions centered around zero with varying dispersion, suggesting moderate uncertainty and partial shrinkage effects.

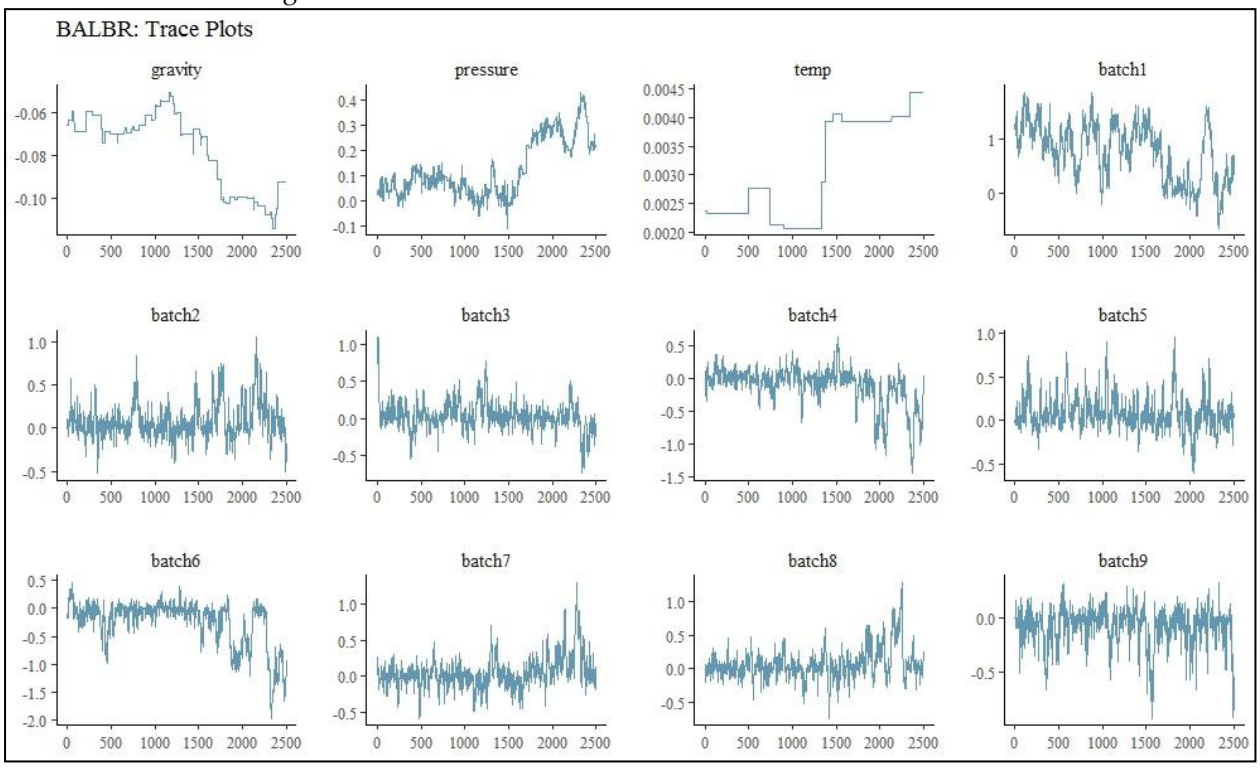

Figure 12. Trace plots of posterior samples for the BALBR model coefficients using the Gasoline Yield dataset

Figure 12 illustrates that the trace plots of the BALBR model reveal stable estimation for the *gravity* and *pressure* coefficients, as evidenced by their good mixing and convergence. In contrast, the *temp* coefficient exhibits a clear shrinkage trend toward zero. Additionally, most *batch* variables fluctuate closely around zero, indicating a stronger shrinkage effect and a sparser model structure compared to the BLBR model.

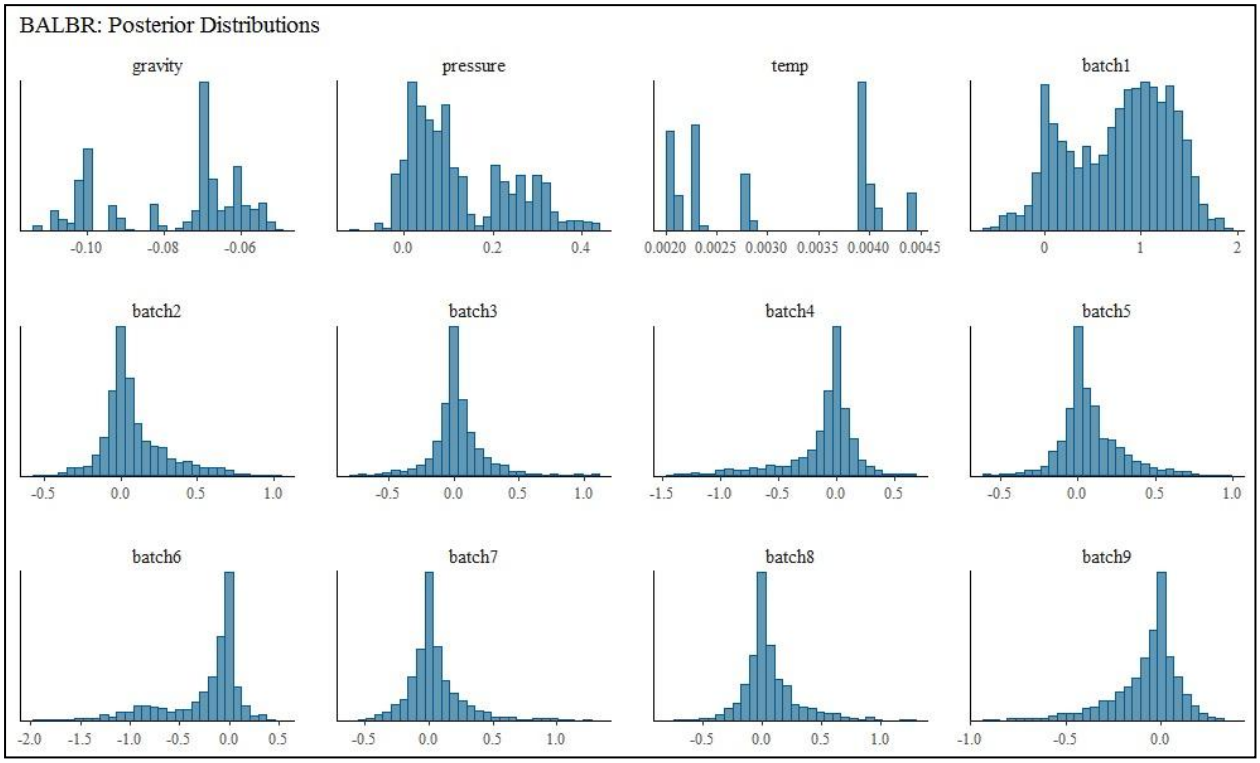

Figure 13. Posterior distributions of the BALBR model coefficients based on the Gasoline Yield dataset

Figure 13 shows that the posterior distributions derived from the BALBR model show that the coefficients for gravity and pressure are distinctly separated from zero, affirming their significance within the model. Conversely, the distribution of temp is tightly concentrated around zero, suggesting strong shrinkage. Additionally, the majority of batch variables display sharply peaked distributions centered at zero, highlighting the model's aggressive shrinkage mechanism and its effectiveness in inducing sparsity in the parameter estimates.

5. Conclusion

This study presented a Bayesian Adaptive Lasso framework for variable selection and parameter estimation within Beta regression models. By incorporating coefficient-specific penalty weights, the proposed model effectively applies adaptive shrinkage, achieving a balance between estimation precision and model sparsity. The simulation results confirmed that the BALBR model consistently outperforms conventional BR, Bayesian BR, and Bayesian Lasso BR approaches in terms of bias reduction, lower standard errors, and improved variable selection accuracy under various conditions. Furthermore, the analysis of real-world data demonstrated the practical utility of the method in distinguishing influential predictors and eliminating irrelevant ones. Overall, the Bayesian Adaptive Lasso Beta Regression model offers a robust and adaptable approach for modeling bounded outcomes, particularly in the presence of high-dimensional predictors.

REFERENCES

[1] S. Al-jajawi and T. Alshaybawee, "Bayesian Variable Selection for Beta Regression

[2] D. F. Andrews and C. L. Mallows, "Scale Mixtures of Normal Distributions," J. R. Stat. Soc. Ser. B (Methodological), vol. 36, no. 1, pp. 99–102, 1974.

[3] J. Fan and J. Lv, "A Selective Overview of Variable Selection in High Dimensional Feature Space," Statistica Sinica, vol. 20, no. 1, pp. 101–148, 2010.

[4] S. Ferrari and F. Cribari-Neto, "Beta Regression for Modelling Rates and Proportions," J. Appl. Stat., vol. 31, no. 7, pp. 799–815, 2004.

[5] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," J. Mach. Learn. Res., vol. 3, pp. 1157–1182, Mar. 2003.

[6] T. Park and G. Casella, "The Bayesian Lasso," J. Am. Stat. Assoc., vol. 103, no. 482, pp. 681–686, 2008.

[7] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," J. R. Stat. Soc. Ser. B (Statistical Methodology), vol. 58, no. 1, pp. 267–288, 1996.

[8] H. Zou, "The Adaptive Lasso and Its Oracle Properties," J. Am. Stat. Assoc., vol. 101, no. 476, pp. 1418–1429, 2006.

[9] Ray Bai, V. Rockova, and E. I. George, "Spike-and-Slab Meets LASSO: A Review of the Spike-and-Slab LASSO," arXiv, Oct. 2020.

[10] C. Leng, M. N. Tran, and D. Nott, "Bayesian Adaptive Lasso," arXiv, Sep. 2010.

[11] Y. Yang, M. J. Wainwright, and M. I. Jordan, "On the Computational Complexity of High-Dimensional Bayesian Variable Selection," arXiv, May 2015.

[12] "Variance Prior Forms for High-Dimensional Bayesian Variable Selection," Bayesian Anal., vol. 14, no. 4, 2019.

[13] "Spike-and-Slab Lasso: Stochastic Search vs. Penalization," in Psychol. Methods, 2023.

[14] "Review of Bayesian Selection Methods for Categorical Predictors," PMCID, 2022.

[15] Y. Wasserman and K. Roeder, "High-Dimensional Variable Selection," arXiv, Apr. 2007.