Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 8 (2026) 3629

Journal of the Nigerian Society of Physical Sciences

Quantile ridge beta regression model and Bayesian regularized quantile beta regression for solving the skewness and improving the accuracy of the model selection in bounded data

Fedaa Noeel Abdulahad[a,b], Majid Khan Majahar Ali[a,*], Raja Aqib Shamim[a,c]

[a] School of Mathematical Sciences, Universiti Sains Malaysia, Gelugor 11800, Pulau Pinang, Malaysia
[b] Department of Mathematics, College of Education for Pure Science, University of Al-Hamdaniya, Bartella 41006, Mosul, Iraq
[c] Department of Mathematics, University of Kotli, Kotli 11100, Azad Jammu and Kashmir, Pakistan

Abstract

Skewed distributions have an additional shape parameter that represents the direction of the asymmetry in the density. If skewness in observations is ignored, statistical inferences based on symmetric distributions may yield biased or even misleading conclusions. Bayesian regularized quantile regression (BRQR) is effective in quantile regression for addressing skewness. Additionally, quantile ridge regression (QRR) has been developed for variable selection and to address multicollinearity among predictor variables. Most studies have used the asymmetric Laplace distribution (ALD) to address skewness, but this solution is not useful for bounded data because it has unbounded support, allowing values far in either direction and potentially violating the bounded-data assumption. The hybrid Bayesian regularized quantile beta regression–quantile ridge beta regression (BRQBR–QRBR) model is proposed for analyzing bounded data within the interval $(0, 1)$ with skewness while improving model-selection accuracy. The proposed model applies BRQBR to the skewed, bounded data to address skewness and then uses QRBR to improve model accuracy. Simulation studies and real-dataset applications were conducted. The results show that the proposed BRQBR–QRBR method, in most cases, outperforms the original method by improving prediction accuracy through skewness correction.

DOI: 10.46481/jnsps.2026.3629

Keywords: Skewed data, Bayesian quantile regression, Beta distribution, Penalized quantile regression, Bounded data

Article History:
Received: 21 June 2026
Received in revised form: 09 August 2026
Accepted for publication: 02 September 2026
Available online: 17 September 2026

Communicated by: P. Thakur

1. Introduction

In many applied settings, the choice of data often yields skewed or heavy-tailed distributions, which are inconsistent with the traditional assumption that the error is normally distributed. Traditional solutions, including variable transformations or robust regression methods, have been proposed to address such deviations. However, skewed distributions can compress important information by encoding asymmetry in the data that symmetric models fail to capture. These distributions include shape parameters that explicitly model the magnitude and direction of skewness Refs. [1–3]; therefore, neglecting skewness may yield biased estimates and flawed inference.

To overcome the shortcomings of classical regression, Koenker and Bassett Ref. [4] proposed quantile regression (QR) as a strong alternative to the ordinary least squares (OLS) method. In contrast to OLS, which focuses only on the conditional mean, QR estimates conditional quantiles, providing a

*Corresponding Author Tel. No.: +60-149-543-405.
Email address: majidkhanmajaharali@usm.my (Majid Khan Majahar Ali)

more detailed picture of the response distribution. By virtue of this feature, QR is especially useful when there is heteroscedasticity, outliers, or distributional asymmetry, and it can also more comprehensively assess the effects of covariates on disparate portions of the outcome distribution Refs. [5, 6].

In recent years, considerable attention has been paid to Bayesian extensions of QR. In Ref. [7], a Bayesian model based on the asymmetric Laplace distribution (ALD) was introduced and the posterior inference was obtained using the Metropolis-Hastings algorithm. Later works by Ref. [8] and Ref. [9] improved the efficiency of estimation using Gibbs sampling schemes. More recently, Ref. [10] proposed a robust Bayesian genomic median regression model for Laplace-distributed errors and outlier-prone data. Despite this, ALD-based Bayesian QR models presuppose unbounded support and are thus not applicable when the response variable is restricted to the unit interval (0,1), since predictions can be out of bounds. Quantile beta regression should be used to provide a better modeling framework in these instances. Recently, Ref. [11] proposed a Bayesian beta regression model with horseshoe priors for high-dimensional bounded responses, which achieved better variable selection and posterior consistency. The study, however, focuses on estimating the coefficients within a single Bayesian framework and does not incorporate posterior predictive information into penalized quantile beta regression.

Proportions, rates, and probabilities are among the bounded responses typical of many fields of study, such as finance, environmental science, public health, and medicine. These data often exhibit skewness, especially when observations are concentrated at the ends of the zero-or-one range. Another trend has been the development of beta regression, first proposed by Ref. [12] which has been a popular method for modeling such data by jointly modeling the conditional mean parameters and the conditional precision using suitable link functions. Generalizations of the beta regression model include the use of covariates in the precision parameter Ref. [13] and the creation of diagnostic statistics to improve model fitting and the variable of interest. Bayesian models of beta regression have also been proposed, including hierarchical models Ref. [14] and mixed beta regression models Ref. [15].

Beta regression models are also prone to multicollinearity among predictors, a problem that inflates variances and compromises estimation, which increases in high-dimensional contexts Refs. [16–18]. To address such difficulties, numerous regularization schemes have been proposed, including ridge-type estimators Ref. [19], Liu-type estimators Refs. [20, 21], James-Stein estimators Ref. [22] and other shrinkage-based schemes Refs. [23, 24]. Similar quantile regression work has conducted parallel penalization, e.g. ridge regression Ref. [25], LASSO Ref. [26], and elastic net Ref. [27], especially with high-dimensional data. More recent efforts have also improved sparse and computationally efficient quantile regression procedures Refs. [28–30], and the use of penalized QR, combined with empirical mode decomposition, to address multicollinearity and non-stationarity in real-world data Refs. [31, 32].

A range of studies have been conducted specifically to characterize skewness and heavy-tailed behavior in bounded datasets. To illustrate, Ref. [33] proposed Bayesian parametric quantile regression based on heavy-tailed distributions, whereas Ref. [34] proposed other unit-distribution-based quantile regression models for educational data. In Ref. [35], a flexible quantile regression model was proposed using a reparametrized generalized beta distribution that simultaneously controls skewness, kurtosis, and tail behavior. Nevertheless, quantile beta regression is frequently not used as a benchmark in comparative studies, which restricts its performance to a direct evaluation.

Despite these advances, significant gaps remain in existing literature. Current models of Bayesian Quantile Beta Regression (BRQBR) primarily address the boundedness and skewness of the response variable using Bayesian shrinkage priors and do not explicitly account for the negative impact of multicollinearity on predictive accuracy. On the other hand, Quantile Ridge Beta Regression (QRBR) stabilizes coefficient estimates even when the regressors are correlated by using ridge penalization, without relying on the predictive information from Bayesian regularization. In the same way, Bayesian shrinkage and penalized quantile regression methods have mainly been developed as separate statistical models to improve prediction performance, rather than as collaborative methods that work together to achieve more accurate predictions.

To overcome the above drawbacks, this study proposes a novel hybrid model: Bayesian Regularized Quantile Beta Regression (BRQBR) combined with Quantile Ridge Beta Regression (QRBR). The proposed approach uses horseshoe priors to model skewness and bounded responses. It simultaneously shrinks estimation noise by fitting a BRQBR model, unlike conventional methods that fit only a linear model within a single framework. The posterior predictive mean from the BRQBR model is then added as an explanatory variable to the QRBR model, following the approach of Ref. [3] and building upon the concepts of Refs. [32, 36]. This predictive augmentation allows the use of information obtained from the Bayesian model while still helping reduce multicollinearity via ridge penalization.

The proposed BRQBR differs from other existing penalized quantile regression methods, such as Bayesian beta regression, Bayesian quantile beta regression, and ridge beta regression, in three important ways. Firstly, it unifies the two procedures, Bayesian regularization and ridge penalization, into a single two-stage modeling approach rather than considering them as independent estimation procedures. Secondly, it casts the posterior predictive augmentation by adding the Bayesian posterior predictive mean as another predictor to the second-stage regression model. Third, it concurrently addresses two important challenges often encountered in bounded data analysis: distributional skewness and multicollinearity among predictors, thereby enhancing predictive accuracy while maintaining the boundedness of the response variable. To our knowledge, we have not seen such a hybrid prediction-augmentation strategy for quantile beta regression models previously proposed.

The methodology is then tested using a comprehensive Monte Carlo simulation study across different degrees of skewness, contamination, and sample sizes, and is subsequently applied to a real environmental problem involving moisture measurements in a solar dryer. The ability of the proposed BRQBR–QRBR framework to predict is evaluated using mean squared error (MSE), mean absolute error (MAE), residual standard deviation, and relative efficiency at various quantile levels.

This paper is organized as follows: Section 2 introduces the Bayesian Regularized Quantile Beta Regression model and the Quantile Ridge Beta Regression model and proposes the BRQBR–QRBR methodology. The simulation study is described in Section 3, and the real data application is presented in Section 4. The cross-validation analysis is reported in Section 5, and the simulation and empirical results are reported in Section 6. The findings are discussed in Section 7, and the limitations and directions for future research are concluded in Section 8.

2. Materials and methods

2.1. Beta regression model

Beta regression is a natural model for analyzing continuous response variables on the unit interval (0,1), including proportions, rates, and probabilities. Where Y is a beta distributed random variable whose shape parameters are a and b where a and b, are greater than 0, and its probability density function is presented by

$$f(y \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, y^{(a-1)}(1-y)^{(b-1)}, \quad 0 < y < 1. \tag{1}$$

Following Ferrari and Cribari-Neto (2004) Ref. [12], the beta distribution is reparametrized in terms of the mean μ and precision parameter ϕ, where

$$\mu = \frac{a}{(a+b)} \quad \text{and} \quad \phi = a + b.$$

Under this parameterization, we write $Y \sim \text{Beta}(\mu, \phi)$.

$$f(y \mid \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}\, y^{(\mu\phi-1)}(1-y)^{(1-\mu)\phi-1}, \quad y \in (0, 1). \tag{2}$$

So, the mean and variance of Y are

$$E(Y) = \mu \quad \text{and} \quad Var(Y) = \frac{\mu(1-\mu)}{(1+\phi)}, \quad 0 < \mu < 1. \tag{3}$$

Let $y_1, ..., y_n$ is independent observations such that $y_i \sim Beta(\mu_i, \phi)$ where the mean parameter is linked to covariates through a logit link,

$$g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta},$$

where $\mathbf{x}_i$ is the i-th row of the design matrix $X \in \mathbb{R}^{n\times p}$ and $\beta \in \mathbb{R}^p$ is a vector of regression coefficients and the precision parameter $\phi > 0$ is assumed constant throughout this study and is modeled separately in the Bayesian framework which is commonly used by Refs. [37, 38].

2.2. Quantile regression framework

Classical linear regression models the conditional mean of the response variable and is sensitive to departures from normality, including skewness and heteroskedasticity. However, quantile regression (QR) was introduced in Ref. [4] as a robust alternative that models conditional quantiles of the response variable. For a linear model

$$y_i = x_i^\top \beta + \varepsilon_i,$$

the τ-th conditional quantile $(0 < \tau < 1)$ of y_i given x_i is

$$Q_{y_i}(\tau \mid x_i) = x_i^\top \beta(\tau).$$

The quantile regression estimator $\hat{\beta}(\tau)$ is obtained by minimizing the check loss function

$$\hat{\beta}(\tau) = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta), \tag{4}$$

where

$$\rho_\tau(u) = \begin{cases} \tau u, & u \geq 0, \\ (\tau - 1)u, & u < 0, \end{cases}. \tag{5}$$

2.3. Quantile beta regression

For bounded responses $y \in (0, 1)$, classical quantile regression models are inappropriate due to their unbounded support however, quantile beta regression addresses this limitation by modeling the conditional distribution of y by using beta distribution. Assume

$$\mu_i = \text{logit}^{-1}(x_i^\top \beta_\tau),$$

and

$$y_i \sim \text{Beta}(\mu_i, \phi).$$

The τ-th conditional quantile of y_i is then given by

$$Q_Y(\tau \mid x_i) = I_\tau^{-1}(\mu_i \phi, (1-\mu_i)\phi), \tag{6}$$

where $I_\tau^{-1}(\cdot)$ denotes the inverse regularized incomplete beta function which allows covariates to directly influence different parts of the conditional distribution while respecting the bounded support and skewness of the data.

2.4. Bayesian regularized quantile beta regression (BRQBR)

To address multicollinearity and overfitting in quantile beta regression, we adopt a Bayesian regularization approach using a global–local shrinkage prior as follows

$$\mu_i = \text{logit}^{-1}(x_i^\top \beta),$$

and

$$y_i \sim \text{Beta}(\mu_i, \phi).$$

The regression coefficients are assigned a horseshoe-type prior Refs. [11, 39], where

$$\begin{aligned} \beta_k \mid \tau, \lambda_k &\sim N(0, \tau^2 \lambda_k^2), \\ \lambda_k &\sim \text{Half-Cauchy}(0, 1), \\ \tau &\sim \text{Half-Cauchy}(0, 1), \end{aligned} \tag{7}$$

where λ_k is a local shrinkage parameter controlling coefficient-specific regularization and τ is a global shrinkage parameter governing overall sparsity. This prior allows strong signals to remain relatively unshrunk while aggressively shrinking weak or noisy predictors toward zero. The precision parameter is assigned a weakly informative prior as

$$\phi \sim \text{Gamma}(2, 1).$$

Posterior inference is conducted using Hamiltonian Monte Carlo as implemented in Stan. The posterior predictive mean

$$\hat{\mu}_i = \mathbb{E}[\mu_i \mid y],$$

is used as a denoised representation of the skewed response.

The local shrinkage parameters λ_k enable the shrinkage of the coefficients of relevant predictors to avoid shrinkage to values that are too small, while the global parameter τ determines the extent of the regularization. The above prior specification is commonly used in Bayesian sparse regression.

For testing purposes, a preliminary sensitivity analysis was performed by performing a scale change to the global horseshoe prior. Also, other priors (with scale 0.5 and 2) were explored, in addition to the default specification τ ˜ Half-Cauchy(0,1). These posterior estimates are very similar across all specifications, as are the predictive performance measures, suggesting that the proposed BRQBR–QRBR framework is relatively robust to moderate changes in prior hyperparameters.

2.5. *Quantile ridge beta regression (QRBR)*

Ridge regression mitigates multicollinearity by penalizing the squared l_2-norm of the regression coefficients, while in the quantile beta regression setting, ridge penalization is applied to the quantile loss on the logit scale. If $z_i = \text{logit}(\mu_i)$ then, quantile ridge beta regression estimator is defined as

$$\hat{\beta}_{\text{QRBR}}(\tau) = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \rho_\tau(z_i - x_i^\top \beta) + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}, \quad (8)$$

where $\lambda \geq 0$ is a tuning parameter selected via cross-validation. This adjustment helps to better estimate the regression parameters more accurately by capturing relationships among the predictor variables Refs. [9, 40].

2.6. *Proposed BRQBR–QRBR method*

To simultaneously address skewness in bounded responses and multicollinearity among predictors, we propose a hybrid approach that integrates Bayesian regularized quantile beta regression with quantile ridge beta regression BRQBR–QRBR as follows.

Step 1: Fit the BRQBR model using the horseshoe prior and obtain posterior predictive means on response $y \in (0, 1)$, predictor matrix $X \in \mathbb{R}^{n\times p}$ with quantile level τ, where

$$\hat{\mu}_i^{\text{BRQBR}} = \mathbb{E}[\mu_i \mid y].$$

Step 2: Construct an augmented predictor matrix as

$$\tilde{X} = [X, \hat{\mu}^{\text{BRQBR}}].$$

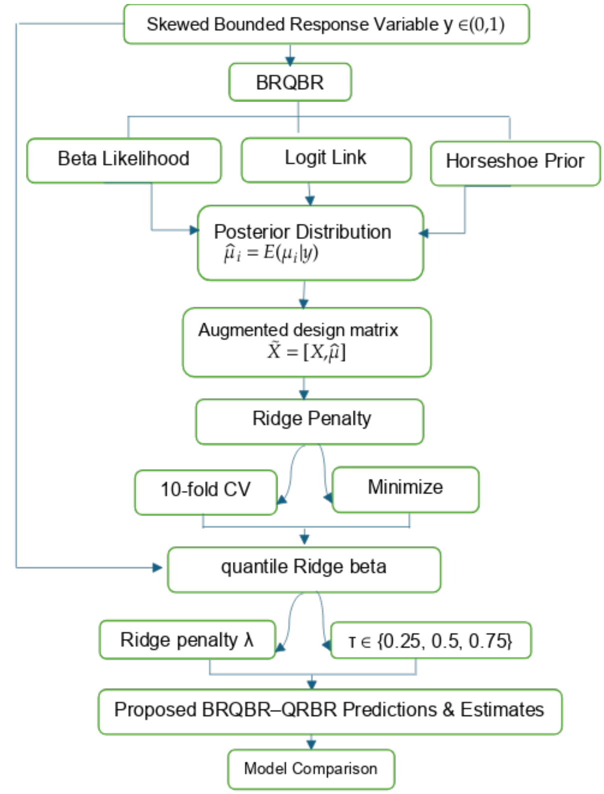

Figure 1: Schematic representation of the proposed BRQBR-QRBR method.

Step 3: Select the ridge tuning parameter λ via 10-fold cross-validation by minimizing the mean squared error.

Step 4: Fit the quantile ridge beta regression model

$$\hat{\beta}_{\text{BRQBR-QRBR}}(\tau) = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \rho_\tau(z_i - \tilde{x}_i^\top \beta) + \lambda \|\beta\|_2^2 \right\}. \quad (9)$$

And get estimated regression coefficients and predicted conditional quantiles.

Step 5: The performance of the proposed BRQBR–QRBR model is evaluated against the original QRBR using predictive accuracy metrics including mean squared error (MSE), mean absolute error (MAE), and residual standard deviation across multiple quantile levels.

The proposed BRQBR–QRBR framework is motivated by the principle that posterior predictive quantities obtained from a Bayesian model provide a statistically efficient summary of the relationship between the response variable and the explanatory variables. Consider the bounded response variable $Y \in (0, 1)$ and the predictor matrix $X = (x_1, \ldots, x_p)$. The data-generating mechanism may be expressed as

$$Y = f(X) + \epsilon,$$

where $f(.)$ denotes an unknown nonlinear regression function and the random error satisfies $E(\epsilon|X) = 0$ and $Var(\epsilon|X) = \sigma^2$. The Bayesian Regularized Quantile Beta Regression (BRQBR)

Table 1: The definition of the variables used in the dataset.

Variable	Definition
Y	Moisture (Response)
H1	Relative Humidity Ambient
H5	Relative Humidity Chamber
PY	Solar Radiation
T1	Temperature (°C) ambient
T2, T3, T4	Temperature (°C) before entering solar collector
T5	Temperature (°C) in front of down v-Groove (Solar Collector)
T6, T8	Temperature (°C) in front of up v-Groove (Solar Collector)
T7, T14, T15, T16, T21, T22	Temperature (°C) Solar Collector
T8, T9, T10, T11, T12	Temperature (°C) behind inside chamber
T13, T17, T18, T19, T23	Temperature (°C) in front of (Inside Chamber)
T20, T23, T24, T25, T28	Temperature (°C) from solar collector to chamber

model estimates the conditional distribution of Y through Bayesian inference with horseshoe shrinkage priors. Let

$$\hat{\mu}_B = E(Y|XD)$$

denote the posterior predictive mean obtained from the fitted BRQBR model, where D represents the observed data. The posterior predictive mean summarizes the systematic information contained in the posterior distribution after accounting for uncertainty in the model parameters. Because the horseshoe prior shrinks weak regression effects toward zero while preserving important signals, $/\hat{mu}_B$ provides a denoised representation of the conditional response structure.

Rather than relying solely on the original predictors, the proposed framework augments the design matrix by incorporating the posterior predictive mean as an additional explanatory variable,

$$\tilde{X} = [X, \hat{\mu}_B].$$

The Quantile Ridge Beta Regression (QRBR) model is then fitted using the augmented predictor matrix,

$$Q_\tau(Y|X) = g^{-1}(X\beta + \gamma\hat{\mu}_B),$$

where $g(.)$ denotes the logit link function, β is the vector of regression coefficients, and γ measures the contribution of the posterior predictive component.

The theoretical motivation for this augmentation is based on bias–variance decomposition. The benefit of Bayesian regularization is that it can shrink unstable coefficients to zero, while retaining those which are important. As a result, the posterior predictive mean includes distributional and nonlinear information that cannot be fully summarized by the original explanatory variables. Adding this amount to the second-stage regression provides more information for prediction while dampening the impact of estimation noise.

Also, the ridge penalty in the QRBR stage addresses multicollinearity by driving correlated regression coefficients toward zero without removing them from the model. As a result, the two stages complement each other. The Bayesian stage is a posterior regularization that provides solid predictions from skewed, bounded data, while the ridge stage provides stability in coefficient estimation when predictors are correlated. This estimator is a hybrid of the two, combining the properties of Bayesian shrinkage and ridge regularization, and has been found to exhibit better predictive stability, especially in cases of highly skewed response variables and small or moderate sample sizes.

Conceptually, this predictive augmentation approach is analogous to stacked prediction, in which predictions from a previous learning stage are used in a later learning stage. The proposed framework, however, uses the posterior predictive mean of a Bayesian quantile beta regression model as an auxiliary predictor in a ridge-penalized quantile beta regression model, in contrast to conventional stacking methods. The integration allows the second-stage model to leverage posterior information about the distribution while maintaining robust ridge estimation in the presence of multicollinearity.

The proposed methodology is a unified approach that accounts for skewness, bounded responses, Bayesian shrinkage, and multicollinearity simultaneously, thereby explaining the increased predictive ability observed in the simulations and empirical studies.

The main steps in applying the proposed BRQBR-QRBR method are summarized in Figure 1.

The Bayesian Regularized Quantile Beta Regression (BRQBR) model was estimated using Hamiltonian Monte Carlo (HMC) in Stan, an open-source probabilistic programming language via the rStan interface. The choice of HMC was inspired by its ability to efficiently navigate high-dimensional posterior distributions, and its convergence is faster and less autocorrelated than that of traditional Markov Chain Monte Carlo (MCMC) methods.

Four independent Markov chains were run simultaneously to ensure reliable posterior inference. The number of iterations per chain was 4,000, with 2,000 for adaptation (warm-up/burn-in) and 2,000 for posterior inference, yielding a total of 8,000 posterior samples across the 4 chains. To reduce divergent transitions and maximize sampling efficiency, an adaptive variant of HMC, called No-U-Turn Sampler (NUTS), was used with a target acceptance probability of 0.95 and a max tree depth of 12.

3. Simulations

The study employed a comprehensive Monte Carlo simulation to assess the performance of the proposed BRQBR-QRBR

Table 2: The clusters with number of variables, Eigen value, percentage of variance and cumulative of percentage of variance for each principal component.

Cluster	No. of Variables	Eigenvalue	% Variance	Cumulative % Variance	Variables
Cluster1	1	–	–	–	H1
Cluster2	1	–	–	–	H5
Cluster3	1	–	–	–	PY
Cluster4	8	7.257528523	90.7191065	90.71911	T1, T2, T3, T4, T6, T13, T23, T29
		0.344690974	4.3086372	95.02774	
		0.215905414	2.6988177	97.72656	
		0.112862069	1.4107759	99.13734	
		0.032375047	0.4046881	99.54203	
		0.021286811	0.2660851	99.80811	
		0.010269674	0.1283709	99.93648	
		0.005081488	0.0635186	100.00000	
Cluster5	17	15.78449	92.84994	92.84994	T5, T8, T14, T15, T16, T21, T22, T9, T10, T11, T12, T17, T19, T25, T26, T27, T28
		0.5436771	3.198101	96.04804	
		0.3223073	1.895925	97.94397	
		0.2491098	1.465352	99.40932	
		0.04169977	0.2452927	99.65461	
		0.02389699	0.1405705	99.79518	
		0.01902288	0.1118993	99.90708	
		0.005773038	0.03395905	99.94104	
		0.005392559	0.03172094	99.97276	
		0.001661366	0.009772743	99.98253	
		0.001147210	0.006748295	99.98928	
		0.000787676	0.004633388	99.99392	
		0.000525899	0.003093521	99.99701	
		0.000369074	0.002171022	99.99918	
		0.000095757	0.000563279	99.99974	
		0.000036131	0.000212537	99.99996	
		0.000007399	0.000043522	100.00000	
Cluster6	1	–	–	–	T7

Table 3: Descriptive statistics for each principal component.

Variable	Mean	SD	Skew.	Kurt.
Y	0.55234	0.1654	0.848239	-0.87574
H1	92.74993	170.1426	2.777199	7.465287
H5	34.77680	17.19086	1.920551	3.823985
PY	330.3889	255.7499	0.564006	-0.78236
PC1-C4	2.7E-12	2.69398	-0.22921	-0.48990
PC1-C5	1.41E-11	3.97297	-0.11656	-0.79015
T7	40.71594	5.22609	1.976757	8.244915

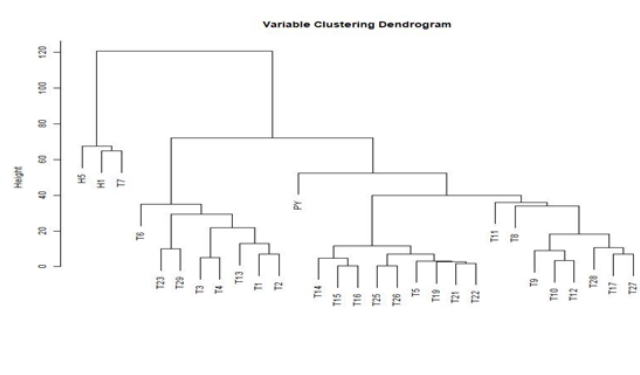

Figure 2: The variable clustering dendrogram.

penalized beta regression models compared with their original counterparts. All simulations and analyses were performed using the R statistical package. The sample sizes of $n = 100$, $n = 200$ and $n = 300$ were used to test small, moderate and large sample behavior. Predictor variables were sampled each time using a multivariate normal distribution with autoregressive correlation structure, $\Sigma_{ij} = \rho^{(|i-j|)}$ with $\rho = 0.7$ and induced moderate covariate multicollinearity. The real regression coefficients were defined as $\beta = (1, 0.5, \ldots, 0.5)/\sqrt{p}$ with $p = 6$ and the mean parameter was associated by a logit function.

To study robustness under various distributional conditions, two bounded data-generating processes were used to produce responses: the Beta distribution with a precision parameter of 10 and the logit-normal distribution. Skew factors of 0.75, 0.95 and 0.999 were used to introduce different levels of skewness

Table 4a: Mean square error (MSE) for the proposed method BRQBR-QRBR and the classical quantile ridge beta regression QRBR for both the beta distribution and the Logit-normal distribution (n = 100). The smallest value indicates better performance.

Distribution	Skew	Out.	MSE-BRQBR-QRBR			MSE-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.015991	0.015064	0.012755	0.015874	0.015264	0.013214
		5%	0.027172	0.026721	0.023596	0.027903	0.027177	0.023758
		10%	0.020588	0.020721	0.018714	0.020271	0.020335	0.01831
	0.95	0%	0.002822	0.002783	0.002599	0.002615	0.002571	0.002414
		5%	0.006116	0.005517	0.004454	0.005557	0.005051	0.004126
		10%	0.003297	0.003343	0.00305	0.002946	0.00306	0.002863
	0.999	0%	0.003034	0.003041	0.002882	0.002447	0.00249	0.002401
		5%	0.00219	0.002262	0.002116	0.002435	0.002523	0.002399
		10%	0.005391	0.005695	0.005456	0.005031	0.005322	0.00511
Logit-n	0.75	0%	5.86E-04	7.33E-04	0.00106	5.70E-04	7.21E-04	0.001045
		5%	0.00129	0.00128	0.001461	0.001194	0.001166	0.001358
		10%	0.022805	0.026	0.026818	0.023067	0.026339	0.027207
	0.95	0%	4.47E-04	3.86E-04	5.51E-04	3.87E-04	3.39E-04	5.13E-04
		5%	0.003178	0.003597	0.003969	0.002623	0.003043	0.00345
		10%	0.00186	0.002005	0.002272	0.00179	0.001943	0.002207
	0.999	0%	6.43E-04	7.22E-04	8.77E-04	5.86E-04	6.78E-04	8.23E-04
		5%	0.003955	0.004425	0.00507	0.003779	0.004254	0.004897
		10%	0.012056	0.010999	0.009681	0.011784	0.010723	0.009424

Table 4b: Same as in Table 4a, but for n = 200.

Distribution	Skew	Out.	MSE-BRQBR-QRBR			MSE-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.020076	0.019332	0.016735	0.020133	0.01947	0.016928
		5%	0.025534	0.024369	0.020895	0.025472	0.024309	0.020847
		10%	0.013844	0.012603	0.010262	0.013945	0.012712	0.010384
	0.95	0%	0.003243	0.002825	0.002206	0.003051	0.002669	0.002108
		5%	0.001074	0.001024	8.98E-04	0.001107	0.001042	9.10E-04
		10%	0.004636	0.005138	0.005117	0.00533	0.005954	0.005996
	0.999	0%	0.001712	0.001648	0.001473	0.00153	0.001465	0.00132
		5%	8.81E-04	9.98E-04	0.001046	0.001037	0.00116	0.001226
		10%	0.001521	0.001603	0.001531	0.001501	0.001567	0.00148
Logit-n	0.75	0%	8.13E-04	5.41E-04	4.59E-04	7.88E-04	5.11E-04	4.23E-04
		5%	0.001828	0.001987	0.002316	0.00187	0.002043	0.002428
		10%	0.005128	0.0064	0.007511	0.005838	0.007147	0.008265
	0.95	0%	7.51E-04	7.73E-04	9.34E-04	7.25E-04	7.55E-04	9.22E-04
		5%	0.002437	0.002033	0.001781	0.003079	0.002814	0.002539
		10%	0.004659	0.003983	0.003719	0.004412	0.003753	0.003507
	0.999	0%	5.61E-04	5.68E-04	7.68E-04	5.48E-04	5.63E-04	7.45E-04
		5%	0.002936	0.002842	0.002874	0.002657	0.002528	0.002553
		10%	0.001913	0.001532	0.001172	0.002096	0.00169	0.001312

Table 4c: Same as in Table 4a, but for n = 300.

Distribution	Skew	Out.	MSE-BRQBR-QRBR			MSE-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.022351	0.02065	0.017138	0.02221	0.020688	0.01733
		5%	0.020474	0.019001	0.01583	0.020572	0.019333	0.016329
		10%	0.009583	0.010289	0.009915	0.009581	0.010341	0.010009
	0.95	0%	0.002072	0.001831	0.001455	0.001868	0.001657	0.001327
		5%	0.002268	0.002199	0.001957	0.00224	0.002147	0.00189
		10%	8.63E-04	8.43E-04	7.41E-04	9.09E-04	8.82E-04	7.68E-04
	0.999	0%	9.99E-04	0.001172	0.001273	9.80E-04	0.001147	0.00125
		5%	0.001171	0.001244	0.001228	0.00121	0.001278	0.001273
		10%	0.001245	0.001276	0.001181	0.00114	0.001177	0.001098
Logit-n	0.75	0%	4.56E-04	3.71E-04	4.93E-04	3.92E-04	3.15E-04	4.35E-04
		5%	0.001153	0.001067	9.37E-04	0.001026	9.09E-04	7.72E-04
		10%	8.35E-04	5.47E-04	5.98E-04	7.93E-04	5.22E-04	5.65E-04
	0.95	0%	2.22E-04	2.46E-04	4.39E-04	1.79E-04	2.06E-04	4.06E-04
		5%	0.001837	0.002087	0.002348	0.001734	0.002006	0.00232
		10%	0.003086	0.004069	0.005044	0.003068	0.004038	0.004973
	0.999	0%	5.00E-04	3.84E-04	3.77E-04	4.66E-04	3.55E-04	3.47E-04
		5%	0.001028	0.001175	0.001564	0.001225	0.001403	0.001798
		10%	8.27E-04	9.00E-04	0.001215	8.29E-04	9.13E-04	0.001236

as in Ref. [3]. Also, contamination cases were included by injecting 0%, 5%, and 10% boundary outliers, thereby testing robustness to extreme outliers.

In each case, the following models were estimated: a classical penalized beta regression using Ridge penalties and its BRQBR-augmented version, in which posterior mean predictions from a Bayesian regularized quantile beta regression (with a horseshoe prior) were added as new covariates. Parameters used in penalization were chosen through 10-fold cross-validation.

Mean squared error (MSE), mean absolute error (MAE) and residual standard deviation (SD) were used to evaluate model performance at quantile levels $(0.25, 0.50, 0.75)$. Relative efficiency: The ratio of the error of the original model to that of the BRQBR -augmented model was calculated; a ratio value exceeding one showed an improvement brought by the proposed method.

4. Real dataset

The current study analyzed the principal factors for clustering. v-Groove data contain hourly solar radiation, temperature, humidity, and moisture content. The data were obtained from the experimental drying process of a seaweed dryer and were optimized for hierarchical clustering. In this paper, a dataset of 1924 observations will be used to study the effect of 29 different independent variables as shown in Table 1. This dataset is used by Ref. [41] Figure 2 shows the variable clustering dendrogram and Table 2 represents the results of the hierarchical clustering of the 29 variables into 6 clusters, including the eigenvalue and the percentage of variance for each principal component within each cluster. The main descriptive statistics were shown in Table 3 and the response variable Y are skewed to the right, which implies that the right tail is long, relative to the left tail.

5. Cross-Validation analysis

We fitted BRQBR-QRBR and QRBR models. We selected three quantiles: $\theta = \{0.25, 0.50, 0.75\}$. To evaluate the predictive ability of the proposed models, we conducted cross-validation. We partitioned the data into training, and testing data. For each environment, we generated 10-fold cross-validation, with 90% of the observations in the training set and 10% in the test set. The idea is to evaluate the model's predictive ability using observations in the testing set. The selection of the appropriate quantile that "best" represents the relationship between phenotypes and predictors is a challenging task Ref. [42], in addition to the best value for ridge parameter λ. Here we propose using the validation set to tune these parameters; this approach is widely used in machine learning Ref. [3]. BRQBR-QRBR and QRBR models were fitted with 90% of observations in the training set, and the remaining 10% in the validation set were used to compute the mean squared error and mean absolute error, which can serve as criteria for selecting between models in cross-validation settings.

6. Results

6.1. Simulation results

The simulation results show that the effectiveness of the proposed BRQBR-QRBR method, compared with the original QRBR, depends on the levels of skewness and contamination, as well as the sample size. In general, the discrepancies between the two methods are relatively small in symmetric or slightly skewed environments with no contamination, where the two methods have nearly the same MSE, MAE, and SD values for quantiles, as shown in Table 4, Table 5, and Table 6. Nevertheless, there are more apparent differences in the case of moderate-to-high skewness and when outliers are available.

In the case of the Beta data-generating mechanism, the suggested method often provides greater efficiency when skewness is intermediate (skew factor = 0.95) and sample sizes are small to medium ($n = 100, 200$), especially at the low and median quantiles. Table 7 indicates that relative efficiency values in such environments can be significantly greater than 1, meaning the prediction error is smaller than in the original QRBR. The effects of contamination are most pronounced below 5% in terms of percentage, indicating that the BRQBR posterior mean should be added as an additional predictor to better handle moderate outliers at the boundary. Tables 4a-6c and Figure 3 show that when skewness is extreme (0.999), the augmented model performs better in certain contamination scenarios but can exhibit a slight efficiency loss when none of the data is contaminated.

For the logit-normal distribution, Tables 4a, 4b, 4c, 5a, 5b, 5c, 6a, 6b and 6c, and Figure 4 also show that the sensitivity of performance differences to contamination levels is higher. At small sample sizes, clean data (0% outliers) QRBR can be a little more effective at times, however. Nevertheless, when contamination is 10% and skewness is moderate, the BRQBR-QRBR method tends to be more stable, with lower MSE and MAE values and reduced variability. These gains are more evident at lower and higher quantiles ($\tau = 0.25$ and 0.75), where skewness and outliers have greater effects.

As anticipated, a larger sample size ($n = 100$ to $n = 300$) reduces the total prediction error of both models, and the relative performance difference also decreases. However, there are a number of moderate skewness-contaminated cases in which the proposed method has a consistent edge, meaning that the BRQBR-QRBR provides additional structural information that increases predictive strength.

Overall, the simulation study evidence suggests that the two methods perform in a similar manner in the ideal environment; however, the suggested BRQBR-QRBR model offers a tangible advantage in the conditions of skew and moderate contamination - the very conditions in which it was created in the first place.

6.2. Real dataset analysis

The newly created BRQBR-QRBR model was tested on data from 1924 hourly measurements of a v-groove solar dryer,

Table 5a: Mean absolute error (MAE) for the proposed method, BRQBR-QRBR, and the classical quantile ridge beta regression, QRBR, for both the beta and Logit-normal distributions ($n = 100$). The smallest value indicates better performance.

Distribution	Skew	Out.	MAE-BRQBR-QRBR			MAE-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.116278	0.113645	0.104124	0.117137	0.115159	0.106188
		5%	0.153423	0.151494	0.140286	0.154003	0.151357	0.139478
		10%	0.129741	0.13	0.122286	0.129029	0.129048	0.121198
	0.95	0%	0.042322	0.042827	0.040974	0.041536	0.042068	0.040339
		5%	0.06048	0.058454	0.052987	0.057923	0.056122	0.051049
		10%	0.045028	0.045599	0.043359	0.042805	0.043683	0.041867
	0.999	0%	0.047816	0.048169	0.045924	0.044074	0.044685	0.04289
		5%	0.03718	0.03815	0.036793	0.03966	0.04079	0.039501
		10%	0.055731	0.057724	0.056219	0.054009	0.055959	0.054544
Logit-n	0.75	0%	0.019034	0.022134	0.027434	0.018803	0.022155	0.027194
		5%	0.029049	0.027394	0.029658	0.027638	0.025993	0.028622
		10%	0.118118	0.124765	0.125491	0.117876	0.124871	0.12588
	0.95	0%	0.016998	0.015347	0.019651	0.015961	0.014605	0.019207
		5%	0.044655	0.049358	0.054069	0.040871	0.045772	0.050834
		10%	0.035146	0.036335	0.039856	0.034752	0.035897	0.039365
	0.999	0%	0.020394	0.021864	0.024614	0.019913	0.021605	0.023774
		5%	0.052116	0.055519	0.056664	0.050936	0.05428	0.055529
		10%	0.08711	0.086934	0.082476	0.08578	0.085739	0.08129

Table 5b: Same as in Table 5a, but for $n = 200$.

Distribution	Skew	Out.	MAE-BRQBR-QRBR			MAE-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.134958	0.132895	0.122776	0.135289	0.133232	0.123119
		5%	0.149898	0.146547	0.134406	0.149801	0.146323	0.134108
		10%	0.103522	0.099643	0.089841	0.104329	0.10045	0.090651
	0.95	0%	0.045201	0.043437	0.039126	0.044135	0.042543	0.038491
		5%	0.026156	0.026226	0.024741	0.026604	0.026566	0.024988
		10%	0.0525	0.055327	0.054739	0.056279	0.059524	0.059129
	0.999	0%	0.032916	0.032975	0.031103	0.030874	0.03094	0.029245
		5%	0.024678	0.026082	0.025937	0.026592	0.028047	0.027901
		10%	0.033458	0.034588	0.033598	0.033042	0.034088	0.033042
Logit-n	0.75	0%	0.021356	0.016965	0.016291	0.021088	0.01663	0.015707
		5%	0.033285	0.035748	0.039433	0.034422	0.036705	0.039636
		10%	0.058791	0.064503	0.067991	0.06313	0.068507	0.071222
	0.95	0%	0.021772	0.021585	0.024528	0.021072	0.021267	0.024228
		5%	0.037494	0.034959	0.033	0.041625	0.040322	0.038992
		10%	0.051731	0.051896	0.049912	0.050412	0.050609	0.048717
	0.999	0%	0.01878	0.018665	0.021883	0.018461	0.018573	0.021441
		5%	0.043622	0.042419	0.042	0.041345	0.040203	0.039741
		10%	0.037313	0.032056	0.026731	0.039949	0.034407	0.028336

Table 5c: Same as in Table 5a, but for $n = 300$.

Distribution	Skew	Out.	MAE-BRQBR-QRBR			MAE-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.13956	0.13507	0.1227	0.139765	0.135606	0.123553
		5%	0.13245	0.12813	0.11619	0.133291	0.129332	0.117666
		10%	0.08824	0.09084	0.08781	0.088273	0.091007	0.088098
	0.95	0%	0.03657	0.03542	0.0322	0.034802	0.03376	0.03076
		5%	0.03946	0.03963	0.03746	0.038975	0.038871	0.036513
		10%	0.02376	0.02381	0.02239	0.02412	0.024101	0.022591
	0.999	0%	0.0265	0.02838	0.02864	0.026271	0.028123	0.028382
		5%	0.02833	0.02937	0.02868	0.028438	0.029479	0.028847
		10%	0.02764	0.02824	0.02712	0.026636	0.027265	0.026229
Logit-n	0.75	0%	0.01712	0.01503	0.01811	0.015888	0.014165	0.016995
		5%	0.02742	0.02746	0.02596	0.025869	0.025236	0.023472
		10%	0.02169	0.01894	0.01864	0.021331	0.018335	0.018241
	0.95	0%	0.0115	0.01202	0.01595	0.010202	0.011013	0.015448
		5%	0.03467	0.0362	0.03827	0.033791	0.035076	0.037766
		10%	0.04586	0.05185	0.05712	0.045752	0.051868	0.057128
	0.999	0%	0.01763	0.01599	0.01538	0.017082	0.01525	0.014572
		5%	0.02249	0.02356	0.02648	0.022454	0.023462	0.026423
		10%	0.02351	0.02645	0.03237	0.025623	0.028316	0.033613

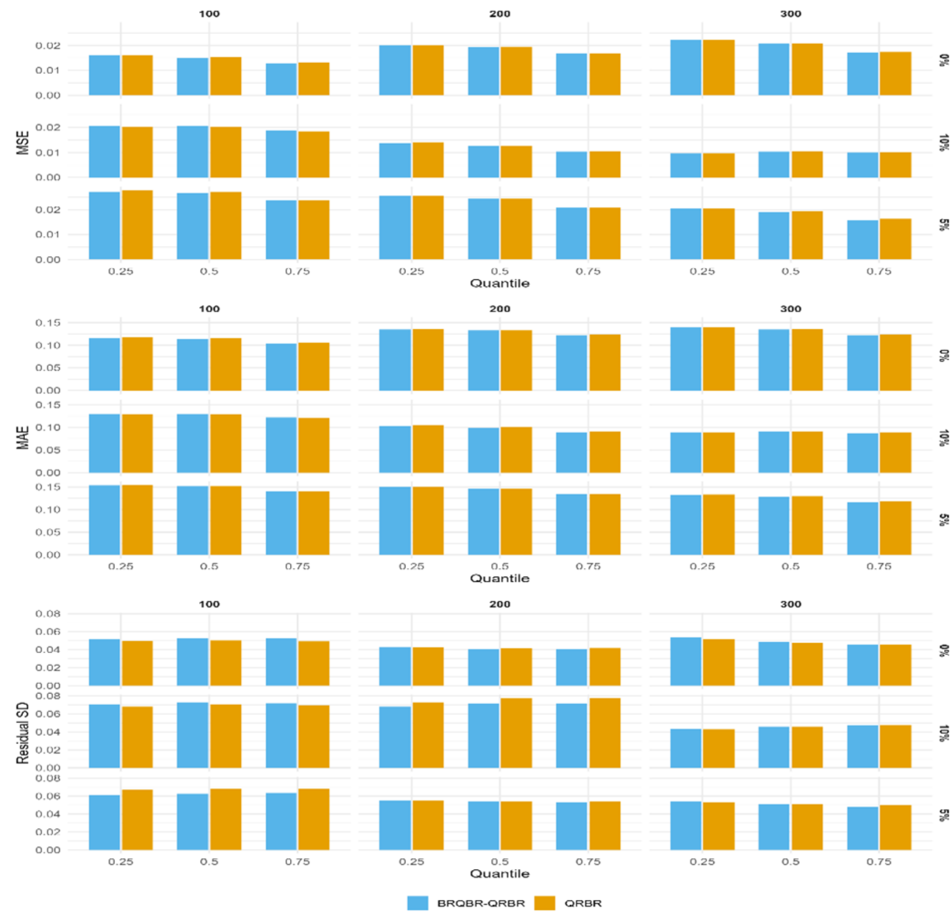

Figure 3: Mean square error (MSE), Mean Absolute Error (MAE) and standard deviation for the proposed method BRQBR-QRBR and classical quantile ridge beta regression QRBR for beta distribution simulated sample.

Table 6a: Standard deviation (SD) for the proposed method BRQBR-QRBR and classical quantile ridge beta regression QRBR for both the beta distribution and the Logit-normal distribution (n = 100). The smallest value indicates better performance.

Distribution	Skew	Out.	SD-BRQBR-QRBR			SD-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.049828	0.046468	0.043855	0.046511	0.044862	0.044139
		5%	0.061328	0.062507	0.063613	0.067256	0.068186	0.068420
		10%	0.062665	0.063055	0.062309	0.061389	0.061754	0.061040
	0.95	0%	0.051860	0.052387	0.051073	0.049800	0.050295	0.049204
		5%	0.061238	0.059383	0.054881	0.056671	0.055057	0.051152
		10%	0.054206	0.054997	0.052967	0.050930	0.052272	0.050967
	0.999	0%	0.051088	0.053129	0.053117	0.045333	0.047600	0.048149
		5%	0.045134	0.046397	0.045329	0.047533	0.049066	0.048403
		10%	0.070622	0.073061	0.071987	0.068074	0.070504	0.069583
Logit-n	0.75	0%	0.022128	0.020487	0.021566	0.021770	0.020305	0.021498
		5%	0.035998	0.034204	0.032883	0.034637	0.032587	0.031374
		10%	0.139046	0.143201	0.140242	0.139966	0.144168	0.141173
	0.95	0%	0.020793	0.015944	0.014744	0.019332	0.014429	0.013234
		5%	0.056202	0.057586	0.056696	0.050928	0.052459	0.051976
		10%	0.043072	0.044376	0.044711	0.042253	0.043657	0.043931
	0.999	0%	0.022946	0.020671	0.019448	0.021658	0.019743	0.018385
		5%	0.062950	0.066066	0.068233	0.061529	0.064770	0.067010
		10%	0.099006	0.099121	0.096720	0.097688	0.097729	0.095363

Table 6b: Same as in Table 6a, but for n = 200.

Distribution	Skew	Out.	SD-BRQBR-QRBR			SD-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.043210	0.040926	0.040811	0.042837	0.041513	0.042124
		5%	0.055429	0.053853	0.053261	0.055130	0.053903	0.053565
		10%	0.057452	0.053681	0.049057	0.057434	0.054180	0.050184
	0.95	0%	0.042909	0.040983	0.037534	0.040955	0.039187	0.036142
		5%	0.027591	0.027602	0.026642	0.028068	0.028091	0.027233
		10%	0.068155	0.071716	0.071520	0.073063	0.077197	0.077440
	0.999	0%	0.038955	0.039346	0.038020	0.036507	0.036863	0.035867
		5%	0.029541	0.030977	0.031036	0.032167	0.033568	0.033749
		10%	0.030465	0.031767	0.031747	0.030086	0.031174	0.030938
Logit-n	0.75	0%	0.027420	0.023278	0.020629	0.026856	0.022600	0.019814
		5%	0.040545	0.044392	0.047937	0.040967	0.045031	0.049041
		10%	0.058822	0.060584	0.062241	0.064414	0.066303	0.067910
	0.95	0%	0.025422	0.027659	0.030338	0.024975	0.027348	0.030108
		5%	0.048732	0.045142	0.041347	0.055039	0.053113	0.049727
		10%	0.057382	0.056989	0.058985	0.055376	0.054956	0.057094
	0.999	0%	0.023491	0.023438	0.024997	0.023209	0.023359	0.024693
		5%	0.053774	0.053250	0.051735	0.051113	0.050216	0.048591
		10%	0.039503	0.036822	0.033570	0.041847	0.038899	0.035552

Figure 4: Mean square error (MSE), Mean Absolute Error (MAE) and standard deviation for the proposed method BRQBR-QRBR and classical quantile ridge beta regression QRBR for Logit-normal distribution simulated sample.

Table 6c: Same as in Table 6a, but for $n = 300$.

Distribution	Skew	Out.	SD-BRQBR-QRBR			SD-QRBR		
			0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.053682	0.049192	0.045833	0.051841	0.048163	0.045771
		5%	0.054180	0.050865	0.048315	0.053066	0.051206	0.050044
		10%	0.043110	0.045747	0.047416	0.042914	0.045897	0.047814
	0.95	0%	0.035674	0.034594	0.032131	0.032846	0.031849	0.029701
		5%	0.040728	0.041427	0.040434	0.040452	0.040993	0.039863
		10%	0.026727	0.026620	0.025218	0.027649	0.027373	0.025716
	0.999	0%	0.031053	0.032757	0.033085	0.030743	0.032376	0.032719
		5%	0.033989	0.035287	0.035004	0.034533	0.035767	0.035642
		10%	0.033344	0.033714	0.032438	0.031766	0.032189	0.031062
Logit-n	0.75	0%	0.020410	0.019275	0.021112	0.018752	0.017752	0.019634
		5%	0.025760	0.027976	0.028746	0.023043	0.025033	0.025820
		10%	0.027740	0.023396	0.022365	0.027056	0.022839	0.021666
	0.95	0%	0.014883	0.014515	0.016725	0.013338	0.013074	0.015617
		5%	0.041474	0.041107	0.039593	0.040331	0.040027	0.038737
		10%	0.051574	0.055032	0.057342	0.051362	0.054726	0.056721
	0.999	0%	0.021299	0.019608	0.018537	0.020497	0.018840	0.017586
		5%	0.028469	0.026964	0.027652	0.028506	0.027164	0.027950
		10%	0.031687	0.034059	0.037373	0.034750	0.037248	0.040496

including variables such as solar radiation, temperature, humidity, and moisture content. The moisture response variable, denoted by Y, exhibits strong right skewness (skewness = 0.848), indicating a pronounced upper tail; quantile-based estimation performs better than mean-based regression, as shown in Table 3. To address severe multicollinearity among the 29 temperature-related covariates, the first step was to perform hierarchical clustering, which yielded six clusters. Essentially all relevant information was retained in the dimensionality reduction process, as the first principal components of the two largest clusters (Cluster 4 and Cluster 5) accounted for 90.72 percent and 92.85 percent of the within-cluster variance, respectively, thereby facilitating stable estimates, as shown in Table 2.

The BRQBR-QRBR methodology was compared with the original QRBR on the 25th, 50th and 75th percentile ($\theta = 0.25, 0.50, 0.75$) in a 10-fold cross-validation scheme (90 percent training, 10 percent testing). Across all studies, the hybrid model consistently had lower mean squared error (MSE) and mean absolute error (MAE) than the reference QRBR. Table 8 and Figure 5 indicate predictive error minimization was the strongest at $\theta = 0.25$, and the relative efficiencies (RE) are above unity in terms of MSE (1.031), MAE (1.005), and standard deviation of the residual value (1.014), indicating predictive accuracy in the lower tail of the moisture distribution is enhanced. When the median was used ($\theta = 0.50$), the gains in performance were small but comparable among measures. The

Table 7: Relative efficiency for the proposed BRQBR-QRBR to the classical QRBR

(a) n = 100

Skew factor	Outlier	Beta			Logit-normal		
		τ = 0.25	0.50	0.75	0.25	0.50	0.75
0.75	0%	0.992654	1.013307	1.035986	0.973087	0.983246	0.985634
	5%	1.026891	1.017061	1.006857	0.925425	0.911338	0.929496
	10%	0.984640	0.981380	0.978436	1.011526	1.013068	1.014511
0.95	0%	0.926526	0.923911	0.928616	0.866587	0.879445	0.930659
	5%	0.908573	0.915658	0.926277	0.825231	0.845976	0.869433
	10%	0.893347	0.915187	0.938799	0.962575	0.968938	0.971181
0.999	0%	0.806464	0.818811	0.833016	0.910345	0.939711	0.938364
	5%	1.111769	1.115401	1.133658	0.955480	0.961430	0.965700
	10%	0.933245	0.934399	0.936695	0.977395	0.974894	0.973434

(b) n = 200

Skew factor	Outlier	Beta			Logit-normal		
		τ = 0.25	0.50	0.75	0.25	0.50	0.75
0.75	0%	1.002856	1.007138	1.011532	0.969561	0.944139	0.921634
	5%	0.997571	0.997523	0.997728	1.022670	1.028090	1.048365
	10%	1.007300	1.008712	1.011867	1.138452	1.116782	1.100507
0.95	0%	0.940845	0.944953	0.955520	0.965972	0.976403	0.986939
	5%	1.030431	1.017221	1.013629	1.263544	1.384379	1.425316
	10%	1.149794	1.158905	1.171875	0.946966	0.942396	0.942891
0.999	0%	0.893657	0.888907	0.895858	0.975874	0.991218	0.969892
	5%	1.177450	1.162427	1.172773	0.905107	0.889546	0.888332
	10%	0.987233	0.977621	0.966473	1.096037	1.103226	1.119580

(c) n = 300

Skew factor	Outlier	Beta			Logit-normal		
		τ = 0.25	0.50	0.75	0.25	0.50	0.75
0.75	0%	0.993696	1.001814	1.011189	0.861483	0.848292	0.881709
	5%	1.004776	1.017454	1.031570	0.889724	0.851933	0.824632
	10%	0.999860	1.005042	1.009459	0.950506	0.953232	0.944251
0.95	0%	0.901523	0.905178	0.911615	0.806553	0.836628	0.925953
	5%	0.987397	0.976060	0.965667	0.944253	0.960851	0.988386
	10%	1.052597	1.046481	1.036298	0.993915	0.992414	0.986105
0.999	0%	0.980995	0.979132	0.981504	0.932106	0.922784	0.919860
	5%	1.033280	1.027535	1.036722	1.002488	1.013735	1.017578
	10%	0.915696	0.922107	0.929639	1.191891	1.194432	1.150072

hybrid model scored better on MSE and MAE with a relatively higher standard deviation of the residual values than the original QRBR; however, this was the case when the quantile was very high (i.e., 0.75).

Taken together, these findings suggest that including the BRQBR-QRBR predicted component as an additional explanatory variable improved predictive stability and efficiency, especially in the lower and upper tails, where skewness exacerbates asymmetry. The moderate scale of improvement, together with the systematic pattern across all measured quantiles, supports the importance of the BRQBR-QRBR framework in the absence of skeletal data for modeling skewed, highly collinear environmental drying data.

7. Discussion

The current study expanded a modeling framework, referred to as BRQBR-QRBR, which includes the predictive component of the Bayesian Regularized Quantile Beta Regression (BRQBR) as an auxiliary predictor within the Quantile Ridge Beta Regression (QRBR) architecture. The main objective was to increase predictive accuracy in limited-response scenarios characterized by skewness, multicollinearity, and outliers. Similar findings have been reported by Refs. [3, 10, 43], though their work relied on the asymmetric Laplace distribution and did not combine Bayesian and ridge-based quantile regression. The performance of the suggested construct was evaluated through an extensive simulation program supplemented by an empirical analysis of solar-dryer moisture measurements.

The simulation study and applications to real data indicate that the proposed hybrid BRQBR–QRBR method consistently outperforms or at least performs competitively with the benchmark methods in predictive performance. The proposed method consistently yielded lower mean squared error (MSE), mean absolute error (MAE), and prediction variability than the

Table 8: Mean square error (MSE) with its relative efficiency, Mean Absolute Error (MAE) with its relative efficiency, and Standard Deviations (SD) with its relative efficiency for the proposed method BRQBR-QRBR and classical quantile ridge beta regression QRBR for a v-groove solar dryer.

Measure	Quantile	BRQBR	QRBR	RE
MSE				
	0.25	0.013245	0.013652	1.030744
	0.50	0.012528	0.012623	1.007532
	0.75	0.013903	0.014038	1.009714
MAE				
	0.25	0.088390	0.088798	1.004616
	0.50	0.089901	0.090530	1.007000
	0.75	0.093758	0.095472	1.018281
SD				
	0.25	0.112953	0.114492	1.013622
	0.50	0.111460	0.111857	1.003528
	0.75	0.113954	0.113677	0.997566

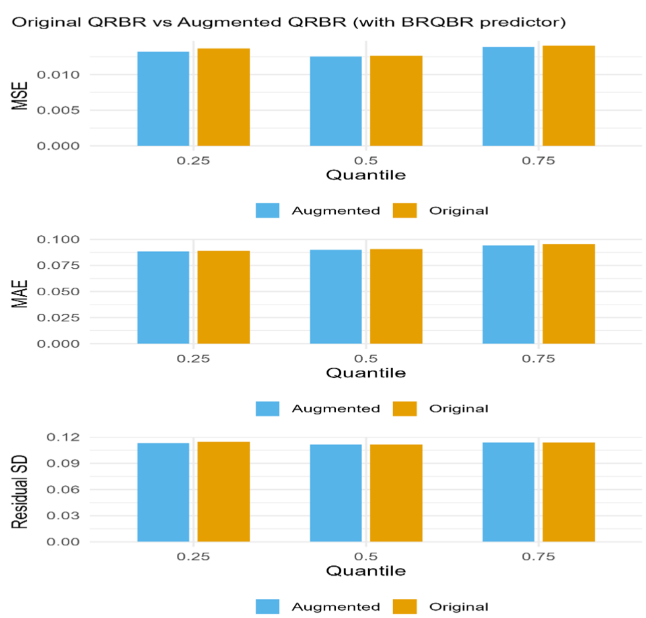

Figure 5: Mean square error (MSE), Mean Absolute Error (MAE) and standard deviation for the proposed method BRQBR-QRBR and classical quantile ridge beta regression QRBR for a v-groove solar dryer.

other models across all sample sizes, response distributions, skewness levels, and outlier contamination settings considered. These enhancements are especially notable in the presence of explanatory variables that are highly collinear and under moderate and high skewness.

The benefits of the proposed framework stem from the complementary strengths of the two stages in the modeling process. In the first stage, the Bayesian Regularized Quantile Beta Regression (BRQBR) model uses horseshoe priors, which enable shrinkage of regression coefficients close to zero while relatively preserving the effects of influential coefficients. This property helps lower the variance of the estimate, stabilize the parameters, and avoid overfitting while preserving valuable predictive information. As a result, the posterior predictive mean from the BRQBR model is a good summary of the distribution of conditional responses, combining the observed data with the posterior uncertainty.

Secondly, the posterior predictive mean is used as an additional predictor in the Ridge-Penalized Quantile Beta Regression (QRBR) model. This augmentation provides a richer predictor space by adding an extra feature that is statistically informed and summarizes the relation captured by this Bayesian estimation, which is complex and nonlinear. The ridge penalty also helps control multicollinearity by reducing the variance of the estimator and keeping all explanatory variables in the model by driving correlated regression coefficients towards the same value. The adaptive Bayesian regularization and ridge penalization procedure is thus more stable and informative than either procedure alone.

As the distribution of responses becomes more skewed, the benefits of the proposed methodology become more apparent. The bounded responses are difficult to model with traditional regression analysis because regression based on the conditional mean may not capture distributional characteristics, especially the tails. To address this, quantile beta regression is proposed, which models conditional quantiles directly, thus offering a richer description of the response distribution. Bayesian regularization is also added, further regularizing the quantile estimates and leading to better predictive performance across the entire range of the response. Likewise, the proposed framework offers clear advantages in the presence of multicollinearity. High correlations among predictors can lead to unstable coefficient estimates, high variance, and low predictive accuracy. Ridge regression is widely recognized as a method for mitigating coefficient instability, and the Bayesian horseshoe prior also provides adaptive regularization to separate important from negligible predictors.

This combination helps the hybrid structure to estimate stably whilst retaining the predictive value of the relevant variables for which the model is valid, which is why it is better when the predictor structures are correlated. The observed changes were not attributable to chance, as formal statistical inference was performed across Monte Carlo replications. The p-value for the comparison of the competing methods in the first simulation scenario was 0.61243, indicating that the difference between the methods was not statistically significant under relatively favorable modeling conditions. The other comparisons, however, yielded p-values of 0.00582 and 0.00068, respectively, indicating that, in more difficult cases, the proposed BRQBR–QRBR approach yielded significantly lower prediction errors than the other approaches. These inferential results corroborate empirical findings from the simulation metrics and indicate that the observed improvement is statistically significant and not attributable to simulation variability.

The method proposed below has practical applications to many applications where bounded continuous responses occur naturally. These range from disease prevalence and treatment success rates in medical research to ecological proportions and species occupancy models, financial risk measures, insurance claim ratios, agricultural experiments, and manufacturing quality-control processes. In these applications, it is sometimes more informative to estimate conditional quantiles than to estimate the conditional mean, as quantiles capture both typical and extreme outcomes. The proposed BRQBR–QRBR framework is thus a flexible and powerful tool for modeling bounded, skewed, and correlated data as well as enhancing predictive performance and estimation stability.

8. Conclusion

This study developed a novel hybrid approach, i.e., Bayesian Regularized Quantile Beta Regression (BRQBR)–Ridge-Penalized Quantile Beta Regression (QRBR), which combines these two methods to model bounded continuous response variables. Proposed methodology: Adaptive Bayesian shrinkage and ridge regularization to address skewness, multicollinearity, and estimation instability, and to enhance predictive ability.

The design of the proposed framework was tested using extensive Monte Carlo simulation experiments and an application to real data. The hybrid method typically outperformed other regression-based techniques in prediction error and estimation stability across a large sample of simulation scenarios with varying sample sizes, response distributions, skewness levels, and outlier contamination. These improvements were also statistically significant in the more challenging simulation settings, as demonstrated by formal statistical analyses, further supporting the robustness of the proposed methodology.

The main strength of this work is that it proves that the posterior predictive mean derived from a Beta regression with Bayesian regularization can be effectively used as an informative auxiliary predictor in the context of the ridge-penalized quantile regression model. This hybrid approach is effective in unifying the uncertainty quantification and the adaptive shrinkage properties of the Bayesian inference with the variance reduction benefits of ridge regularization and hence yields better prediction results for bounded and highly skewed response variables.

While these are positive findings, there are some caveats. In the present study, a single Bayesian shrinkage prior was considered, and ridge penalization in the second-stage model was investigated. In addition, moderate-dimensional predictor settings and independent observations have been studied in the simulation. Future studies should explore other shrinkage priors within the Bayesian framework, adaptive or data-driven penalty choices, elastic-net and non-convex penalization, nonlinear regression, problems with more predictors than observations, and extensions to longitudinal, mixed-effects, spatial, and zero-one-inflated beta regression models. Further theoretical development of the consistency and asymptotic properties of the proposed hybrid estimator would enhance its methodological foundation.

In general, the proposed BRQBR- QRBR framework offers a flexible, powerful, and numerically affordable method for quantile regression with bounded continuous responses. Empirical and inferential results suggest that the method may provide a viable alternative to predictive modeling in applications with a skewed response distribution, multidimensional variable distributions, and nonindependent variables.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declaration of competing interest

The authors declare no conflicts of interest.

Funding

This research received no external funding.

References

[1] S. M. Ahmed, M. K. M. Ali & A. H. Hasan, "Evaluating feature selection methods in a hybrid Weibull Freund-Cox proportional hazards model for renal cell carcinoma", Nigerian Society of Physical Sciences **7** (2025) 2812. https://doi.org/10.46481/jnsps.2025.2812.

[2] Q. Li, N. Lin & R. Xi, "Bayesian regularized quantile regression", Bayesian Analysis **5** (2010) 533. https://doi.org/10.1214/10-BA521.

[3] P. Pérez-Rodríguez, O. A. Montesinos-López, A. Montesinos-López & J. Crossa, "Bayesian regularized quantile regression: A robust alternative for genome-based prediction of skewed data", The Crop Journal **8** (2020) 713. https://doi.org/10.1016/j.cj.2020.04.009.

[4] R. Koenker & G. Bassett Jr., "Regression quantiles", Econometrica **46** (1978) 33. https://doi.org/10.2307/1913643.

[5] Y. Tian & X. Song, "Bayesian bridge-randomized penalized quantile regression", Computational Statistics & Data Analysis **144** (2020) 106876. https://doi.org/10.1016/j.csda.2019.106876.

[6] Y. Wu & Y. Liu, "Variable selection in quantile regression", Statistica Sinica **19** (2009) 801. https://www.jstor.org/stable/24308857.

[7] K. Yu & R. A. Moyeed, "Bayesian quantile regression", Statistics & Probability Letters **54** (2001) 437. https://doi.org/10.1016/S0167-7152(01)00124-9.

[8] E. G. Tsionas, "Bayesian quantile inference", Journal of Statistical Computation and Simulation **73** (2003) 659. https://doi.org/10.1080/0094965031000064463.

[9] H. Kozumi & G. Kobayashi, "Gibbs sampling methods for Bayesian quantile regression", Journal of Statistical Computation and Simulation **81** (2011) 1565. https://doi.org/10.1080/00949655.2010.496117.

[10] A. Montesinos-Lopez, O. A. Montesinos-Lopez, E. R. Villa-Diharce, D. Gianola & J. Crossa, "A robust Bayesian genome-based median regression model", Theoretical and Applied Genetics **132** (2019) 1587. https://doi.org/10.1007/s00122-019-03303-6.

[11] T. T. Mai, "Handling bounded response in high dimensions: A horseshoe prior Bayesian beta regression approach", Bayesian Analysis (2026) 1. https://arxiv.org/abs/2505.22211.

[12] S. Ferrari & F. Cribari-Neto, "Beta regression for modelling rates and proportions", Journal of Applied Statistics **31** (2004) 799. https://doi.org/10.1080/0266476042000214501.

[13] M. Smithson & J. Verkuilen, "A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables", Psychological Methods **11** (2006) 54. https://doi.org/10.1037/1082-989X.11.1.54.

[14] A. J. Branscum, W. O. Johnson & M. C. Thurmond, "Bayesian beta regression: Applications to household expenditure data and genetic distance between foot-and-mouth disease viruses", Australian & New Zealand Journal of Statistics **49** (2007) 287. https://doi.org/10.1111/j.1467-842X.2007.00481.x.

[15] J. I. Figueroa-Zúñiga, R. B. Arellano-Valle & S. L. P. Ferrari, "Mixed beta regression: A Bayesian perspective", Computational Statistics & Data Analysis **61** (2013) 137. https://doi.org/10.1016/j.csda.2012.12.002.

[16] G. James, D. Witten, T. Hastie & R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, 1st ed., Springer, New York, USA, 2013. https://doi.org/10.1007/978-1-4614-7138-7.

[17] A. Junaid, A. Khan, A. Alrumayh, F. M. Alghamdi, E. Hussam, H. M. Aljohani & A. Alrashidi, "Modified two parameter ridge estimator for beta regression model", Journal of Radiation Research and Applied Sciences **17** (2024) 100905. https://doi.org/10.1016/j.jrras.2024.100905.

[18] M. Qasim, K. Månsson & B. M. Golam Kibria, "On some beta ridge regression estimators: Method, simulation and application", Journal of Statistical Computation and Simulation **91** (2021) 1699. https://doi.org/10.1080/00949655.2020.1867549.

[19] M. N. Akram, M. Amin, A. Elhassanein & M. A. Ullah, "A new modified ridge-type estimator for the beta regression model: Simulation and application", AIMS Mathematics **7** (2022) 1035. https://www.aimspress.com/article/doi/10.3934/math.2022062.

[20] A. Erkoç, E. Ertan, Z. Y. Algamal & K. U. Akay, "The beta Liu-type estimator: Simulation and application", Hacettepe Journal of Mathematics and Statistics **52** (2023) 828. https://doi.org/10.15672/hujms.1145607.

[21] P. Karlsson, K. Månsson & B. M. Golam Kibria, "A Liu estimator for the beta regression model and its application to chemical data", Journal of Chemometrics **34** (2020) e3300. https://doi.org/10.1002/cem.3300.

[22] S. Seifollahi & H. Bevrani, "James-Stein type estimators in beta regression model: Simulation and application", Hacettepe Journal of Mathematics and Statistics **52** (2023) 1046. https://doi.org/10.15672/hujms.1122207.

[23] M. R. Abonazel, Z. Y. Algamal, F. A. Awwad & I. M. Taha, "A new two-parameter estimator for beta regression model: Method, simulation, and application", Frontiers in Applied Mathematics and Statistics **7** (2022) 780322. https://doi.org/10.3389/fams.2021.780322.

[24] M. R. Abonazel, I. Dawoud, F. A. Awwad & A. F. Lukman, "Dawoud–Kibria estimator for beta regression model: Simulation and application", Frontiers in Applied Mathematics and Statistics **8** (2022) 775068. https://doi.org/10.3389/fams.2022.775068.

[25] A. E. Hoerl & R. W. Kennard, "Ridge regression: Applications to nonorthogonal problems", Technometrics **12** (1970) 69. https://doi.org/10.1080/00401706.1970.10488635.

[26] R. Tibshirani, "Regression shrinkage and selection via the lasso", Journal of the Royal Statistical Society Series B: Statistical Methodology **58** (1996) 267. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[27] H. Zou & T. Hastie, "Regularization and variable selection via the elastic net", Journal of the Royal Statistical Society Series B: Statistical Methodology **67** (2005) 301. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[28] Q. Tang, Y. Gu & B. Wang, "fastkqr: A fast algorithm for kernel quantile regression", Journal of Computational and Graphical Statistics **35** (2026) 395. https://doi.org/10.1080/10618600.2025.2541004.

[29] K. M. Tan, L. Wang & W.-X. Zhou, "High-dimensional quantile regression: Convolution smoothing and concave regularization", Journal of the Royal Statistical Society Series B: Statistical Methodology **84** (2022) 205. https://doi.org/10.1111/rssb.12485.

[30] N. H. A. Afouna & M. K. M. Ali, "Optimizing precision farming: Enhancing machine learning efficiency with robust regression techniques in high-dimensional data", Nigerian Society of Physical Sciences **7** (2025) 2314. https://doi.org/10.46481/jnsps.2025.2314.

[31] A. R. M. Alsayed, "Turkish stock market from pandemic to Russian invasion, evidence from developed machine learning algorithm", Computational Economics **62** (2023) 1107. https://doi.org/10.1007/s10614-022-10293-z.

[32] A. S. A. Ambark & M. T. Ismail, "Penalized quantile regression and empirical mode decomposition for improving the accuracy of the model selection", Pakistan Journal of Statistics **40** (2024) 199. https://www.pakjs.com/wp-content/uploads/2024/04/40204.pdf.

[33] E. S. B. de Oliveira, M. de Castro, C. L. Bayes & J. L. Bazan, "Bayesian quantile regression models for heavy tailed bounded variables using the No-U-Turn sampler", Computational Statistics **40** (2025) 3007. https://doi.org/10.1007/s00180-022-01297-2.

[34] M. Ç. Korkmaz, C. Chesneau & Z. S. Korkmaz, "A new alternative quantile regression model for the bounded response with educational measurements applications of OECD countries", Journal of Applied Statistics **50** (2023) 131. https://doi.org/10.1080/02664763.2021.1981834.

[35] M. Bourguignon, D. I. Gallardo & H. Saulo, "Parametric quantile beta regression model", International Statistical Review **92** (2024) 106. https://doi.org/10.1111/insr.12564.

[36] A. S. A. Ambark, M. T. Ismail, A. S. Al-Jawarneh & S. A. A. Karim, "Elastic net penalized quantile regression model and empirical mode decomposition for improving the accuracy of the model selection", IEEE Access **11** (2023) 26152. https://doi.org/10.1109/access.2023.3257032.

[37] S. L. P. Ferrari, P. L. Espinheira & F. Cribari-Neto, "Diagnostic tools in beta regression with varying dispersion", Statistica Neerlandica **65** (2011) 337. https://doi.org/10.1111/j.1467-9574.2011.00488.x.

[38] A. B. Simas, W. Barreto-Souza & A. V. Rocha, "Improved estimators for a general class of beta regression models", Computational Statistics & Data Analysis **54** (2010) 348. https://doi.org/10.1016/j.csda.2009.08.017.

[39] J. Piironen & A. Vehtari, "On the hyperprior choice for the global shrinkage parameter in the horseshoe prior", Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, (2017) 905.https://doi.org/10.48550/arXiv.1610.05559.

[40] A. Bager, M. Roman, M. Algedih & B. Mohammed, *Addressing multicollinearity in regression models: A ridge regression application*, MPRA Paper 81390, University Library of Munich, Germany, 2017. https://mpra.ub.uni-muenchen.de/81390/.

[41] M. E. Suhaeri, Alimudin, A. Javaid, M. T. Ismail & M. K. M. Ali, "Evaluation of clustering approach with Euclidean and Manhattan distance for outlier detection", AIP Conference Proceedings **2423** (2021) 070025. https://doi.org/10.1063/5.0075570.

[42] M. Nascimento, F. F. e Silva, M. D. V. de Resende, C. D. Cruz, A. C. C. Nascimento, J. M. S. Viana, C. F. Azevedo & L. M. A. Barroso, "Regularized quantile regression applied to genome-enabled prediction of quantitative traits", Genetics and Molecular Research **16** (2017) gmr16019538. https://pubmed.ncbi.nlm.nih.gov/28340274/.

[43] D. Gianola, A. Cecchinato, H. Naya & C.-C. Schön, "Prediction of complex traits: Robust alternatives to best linear unbiased prediction", Frontiers in Genetics **9** (2018) 195. https://doi.org/10.3389/fgene.2018.00195.