

Bayesian Regularized Quantile Beta Regression for Robust Estimation in Skewed Bounded Data

(Regresi Beta Kuantil Teratur Bayesian untuk Penganggaran Teguh dalam Data Pencong Terbatas)

FEDAA NOEEL ABDULAHAD^{1,2}, MAJID KHAN MAJAHAR ALI^{2,*} & ALAA ADNAN³

¹*University of Al Hamdaniya, College of Education, Department of Mathematics, Mosul-Iraq*

²*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia*

³*University of Glasgow, James Watt School of Engineering, Glasgow, United Kingdom*

Received: 26 November 2025/Accepted: 17 August 2026

ABSTRACT

Skewed distributions often contain shape parameters that determine the direction and magnitude of asymmetry. In other cases, skewness arises naturally from the form of the distribution. Ignoring skewness when modeling with symmetric distributions may yield biased or misleading inferences. Bayesian regularized quantile regression has proven effective for skewed responses, yet existing approaches rely on the asymmetric Laplace distribution (ALD), whose unbounded support makes it unsuitable for bounded data. To address this limitation, we propose a Bayesian Regularized Quantile Beta Regression (BRQBR) model for analyzing bounded data supported on (0,1) with inherent skewness. The proposed model estimates conditional quantiles of a Beta-distributed response using a hierarchical Bayesian regularization framework with global-local shrinkage priors. A Gibbs sampler is developed for posterior computation, and the model's performance is evaluated under different skewness levels and contamination scenarios (5% and 10% outliers) using Beta and logit-normal distributions. Application to a real-world seaweed drying dataset demonstrates consistent improvements in predictive accuracy. Across simulation and empirical analysis, BRQBR outperforms or matches maximum likelihood estimation (MLE) while exhibiting strong robustness to outliers. The proposed framework offers a flexible and accurate solution for modeling skewed bounded responses.

Keywords: Bayesian quantile regression; Beta distribution; robust regression; skewness

ABSTRAK

Taburan pencong selalunya mengandungi parameter bentuk yang menentukan arah dan magnitud asimetri. Dalam kes lain, kepencongan timbul secara semula jadi daripada bentuk taburan. Mengabaikan kepencongan semasa pemodelan dengan taburan simetri boleh menghasilkan inferens yang berat sebelah atau mengelirukan. Regresi kuantil terlaras Bayesian telah terbukti berkesan untuk tindak balas pencong, namun pendekatan sedia ada bergantung pada taburan Laplace asimetri (ALD), yang sokongan tidak terbatasnya menjadikan ia tidak sesuai untuk data terbatas. Untuk menangani batasan ini, kami mencadangkan model Regresi Beta Kuantil Teratur Bayesian (BRQBR) untuk menganalisis data terbatas yang disokong pada (0,1) dengan kepencongan yang wujud. Model yang dicadangkan menganggarkan kuantil bersyarat bagi tindak balas teragih Beta menggunakan rangka kerja teratur Bayesian berhierarki dengan prior pengecutan global-tempatan. Pensampel Gibbs dibangunkan untuk pengiraan posterior dan prestasi model dinilai di bawah tahap kepencongan dan senario pencemaran yang berbeza (5% dan 10% outlier) menggunakan taburan Beta dan logit-normal. Aplikasi pada set data pengeringan rumpai laut dunia sebenar menunjukkan peningkatan yang tekak dalam ketepatan ramalan. Merentasi simulasi dan analisis empirik, BRQBR mengatasi atau memadankan anggaran kemungkinan maksimum (MLE) sambil menunjukkan keteguhan yang kuat kepada penciran. Rangka kerja yang dicadangkan menawarkan penyelesaian yang fleksibel dan tepat untuk memodelkan tindak balas pencong terbatas.

Kata kunci: Kepencongan; regresi kuantil bayesian; regresi teguh; taburan beta

INTRODUCTION

Many real-world datasets exhibit skewed or heavy-tailed distributions, which violate the normality assumptions commonly required in classical regression. Transformations or robust regression are often used to mitigate this issue, but they may distort interpretation or perform poorly when

data are bounded. Skewed distributions are important in applied modeling because they capture asymmetric patterns that arise naturally in economics, environmental science, biology, and engineering (Li, Xi & Lin 2010; Pérez-Rodríguez et al. 2022). When skewness is ignored, estimation and inference become unreliable.

Quantile regression (QR), introduced by Koenker and Bassett (1978), provides a flexible framework for characterizing the conditional distribution of a response variable beyond its mean. It is widely used for heterogeneous data and is more robust than least squares regression has received considerable attention since the seminal work of Yu and Moyeed (2001), who used the asymmetric Laplace distribution (ALD) as a working likelihood. Later extensions improved computational methods and robustness (Montesinos-López et al. 2019; Tsionas 2003).

However, the ALD-based Bayesian QR framework is unsuitable for bounded responses such as proportions, probabilities or rates because the ALD has unbounded support. This mismatch may yield predictions outside the (0,1) interval, violating the nature of the data. For bounded outcomes, the Beta distribution, introduced in a regression framework by Ferrari and Cribari-Neto (2004), is a natural choice because it accommodates asymmetry and heteroscedasticity.

Recent advances in Beta regression have addressed issues such as varying dispersion (Espinheira, Ferrari & Cribari-Neto 2008; Smithson & Verkuilen 2006) and multicollinearity (Abonazel et al. 2022; Akram et al. 2022; Karlsson, Månsson & Kibria 2020). Nevertheless, existing Beta regression approaches are primarily mean-based and are not tailored to quantile estimation, outlier resistance or strong skewness.

To overcome these limitations, we propose a Bayesian Regularized Quantile Beta Regression (BRQBR) model that extends existing quantile Beta regression frameworks by incorporating hierarchical regularization via global-local shrinkage priors. This contributes a robust, flexible method for bounded skewed data subject to outliers and multicollinearity. The main contributions of this work are: 1) Develop a Quantile Beta Regression model capable of estimating conditional quantiles of bounded responses, 2) Construct a Bayesian regularization framework based on global-local shrinkage priors for robust estimation and variable selection, and 3) Evaluate the robustness of the proposed method under skewness and contamination through simulation studies and real-data applications.

The remainder of the paper is organized as follows: First, Beta regression fundamentals are reviewed. Next, the proposed Bayesian Regularized Quantile Beta Regression (BRQBR) model and its Bayesian formulation are introduced. Subsequently, simulation results are presented, followed by an application to a real dataset. Finally, the results are discussed, and the main conclusions are presented.

MATERIALS AND METHODS

BETA REGRESSION MODEL

Beta distribution is a natural choice for modelling continuous data restricted to the interval (0, 1). The density function of the Beta distribution is

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

where $a, b > 0$. Following Ferrari and Cribari-Neto (2004), we reparametrize the distribution in terms of mean μ and precision ϕ . The mean and variance of Y are, respectively:

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

Under the mean-precision parameterization of Ferrari and Cribari-Neto (2004), the probability density function of the Beta distribution can be expressed as $Y \sim B(\mu, \phi)$ where:

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

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

$\phi > 0$, denotes the precision parameter. Let $y_1, y_2, \dots, y_n$ be independent random variables from $Beta(\mu_i, \phi_i)$, $i = 1, 2, \dots, n$. A link function of the mean μ_i can be written as $g(\mu_i) = \eta_i = x_i^T \beta$, and a link function of the precision ϕ can be expressed as $h(\phi) = \zeta_i = z_i^T \gamma$ where x_i and z_i are the i^{th} row of the design matrix X and Z of order $n \times p$ and $n \times q$, respectively, $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_q)^T$ are vectors of unknown parameters of orders $p \times 1$ and $q \times 1$, respectively. The link functions $g(\cdot)$ and $h(\cdot)$ are strictly monotonic and twice differentiable such that $g(\cdot) : (0, 1) \rightarrow \mathbb{R}$ and $h(\cdot) : (0, \infty) \rightarrow \mathbb{R}$. The link functions are the logit link $\mu(1-\mu)$ for the mean parameter and the log link $h(\phi) = \log(\phi)$ for the precision parameter commonly used by Ferrari, Espinheira and Cribari-Neto (2011) and Simas, Barreto-Souza and Rocha (2010).

BAYESIAN REGULARIZED QUANTILE BETA REGRESSION MODEL

QUANTILE REGRESSION MODEL

Consider the following standard linear regression model:

$$y_i = x_i^T \beta + w_i$$

where $x_i^T \beta$ represents the conditional mean of y_i given the covariate vector x_i and w_i denotes the random error term. Unlike ordinary least squares regression, quantile regression does not require a specific distributional assumption for w_i . It is only assumed that the conditional τ -th quantile of w_i equals zero. The τ -th ($0 < \tau < 1$) quantile of w_i is the value, q_τ , for which $p(w_i \leq q_\tau) = \tau$. The τ -th conditional quantile of y_i given x_i is defined as $q_\tau(y_i|x_i) = x_i^T \beta(\tau)$ where $\beta(\tau)$ is a vector of coefficients

dependent on τ . And it is defined as any solution, $\beta(\tau)$ to the following quantile regression minimization problem (Yu & Moyeed 2001):

$$\hat{\beta}_\tau = \arg \min_{\beta} \sum_{i=1}^n \rho_\tau(y_i - x_i^T \beta) \quad (4)$$

where $\tau_0 = \rho(u)$ is the check function (loss function), defined as (Pérez-Rodríguez et al. 2022):

$$\rho_\tau(u) = u(\tau - I_{(u < 0)}) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (1 - \tau)u & \text{if } u < 0 \end{cases} \quad (5)$$

The loss function is widely used in quantile regression because minimizing it yields estimates of conditional quantiles. Compared with squared-error loss, it is less sensitive to extreme observations and therefore provides improved robustness in the presence of skewness and outliers (Koenker & Bassett 1978).

QUANTILE BETA REGRESSION MODEL

We model the conditional distribution of $y \in (0,1)$ by a Beta distribution parameterized by a mean $\mu(x)$ and precision ϕ and $\mu_i = \text{logit}^{-1}(x_i^T \beta_\tau)$. The τ -th quantile is given by the inverse regularized incomplete Beta function:

$$Q_\tau(\tau|x) = I_\tau^{-1}(\mu(x)\phi, (1 - \mu(x))\phi) \quad (6)$$

where I_τ^{-1} is the inverse of the regularized incomplete Beta function. This formulation provides a parametric mapping from covariates to conditional quantiles and is particularly useful when the response is restricted to the interval (0,1) and exhibits heteroscedasticity and skewness. The parameters of the quantile Beta regression model are estimated using maximum likelihood estimation (MLE), for which the cumulative distribution function for Beta distribution is $F(y|\mu, \phi) = I_y(\mu\phi, (1 - \mu)\phi)$ the regularized incomplete Beta function.

The likelihood function is

$$L(\beta, \gamma, \phi|y) = \prod_{i=1}^n \frac{\Gamma(\phi)}{\Gamma(\mu_i\phi)\Gamma((1-\mu_i)\phi)} y_i^{\mu_i\phi-1} (1-y_i)^{(1-\mu_i)\phi-1} \quad (7)$$

where $\mu_i = g^{-1}(x_i^T \beta)$, $\phi_i = h^{-1}(z_i^T \gamma)$.

Equation (7) represents the likelihood formulation adopted in the proposed BRQBR framework. While it is based on the standard Beta likelihood, its integration with the Bayesian regularization structure and quantile-based modeling strategy constitutes a novel contribution of the present study.

PRIOR DISTRIBUTIONS

To induce shrinkage and improve robustness against overfitting, we adopted a hierarchical global-local prior structure for the regression coefficients β . This formulation has been inspired by the Bayesian regularized quantile regression framework of Pérez-Rodríguez et al. (2022), where it introduces adaptive regularization that allows strong coefficients to remain unshrunk while shrinking weak effects toward zero. Specifically, we employed a horseshoe-type prior defined as follows (Piironen & Vehtari 2017):

$$\begin{aligned} \beta_k | \tau, \lambda_k &\sim N(0, \tau^2 \lambda_k^2), \\ \lambda_k &\sim \text{Half-Cauchy}(0, 1), \\ \tau &\sim \text{Half-Cauchy}(0, 1), \end{aligned} \quad (8)$$

where λ_k is a local shrinkage parameter controlling the amount of regularization applied to the j -th coefficient, and τ is a global shrinkage parameter that determines the overall degree of regularization across all predictors. The half-Cauchy prior possesses heavy tails, allowing important coefficients to escape excessive shrinkage while strongly shrinking noise coefficients toward zero. The hierarchical structure (8) acts as a continuous analogue of LASSO and ridge penalties within a fully Bayesian framework and yields automatic data-driven regularization (Carvalho, Polson & Scott 2010; Pérez-Rodríguez et al. 2022).

For the precision parameter ϕ of the Beta distribution, we used a weakly informative Gamma prior, $\phi \sim \text{Gamma}(2, 1)$ which ensures positive support and moderate prior concentration around plausible dispersion values. All priors were selected to balance model flexibility and stability while mitigating overfitting, following recommendations from contemporary Bayesian hierarchical modeling and shrinkage-prior literature (Gelman et al. 2013; McElreath 2020; Piironen & Vehtari 2017). Convergence and posterior sampling quality were assessed using modern Bayesian diagnostic principles (Vehtari et al. 2021).

POSTERIOR DISTRIBUTION

The joint posterior distribution of all unknown quantities can be obtained by applying Bayes theorem, so by combining Equation (7) and Equation (8) we obtain:

$$p(\beta, \gamma, \phi, \lambda, \tau | y) \propto L(\beta, \gamma, \phi | y) \prod_{j=1}^p N(\beta_j | 0, \lambda_j^2 \tau^2) HC(\lambda_j) HC(\tau) p(\gamma) p(\phi) \quad (9)$$

The joint posterior distribution is too complex to be handled directly in closed form. This complexity arises because the joint distribution involves multiple

parameters and their interactions, making it difficult to derive direct analytical expressions or perform calculations. So, hierarchical structures can simplify the problem by breaking it down into layers of conditional relationships. This allows us to focus on each parameter's distribution given the others, which can be more manageable to implement a Gibbs sampler (Yu & Moyeed 2001).

SOFTWARE

The BRQBR model was fitted using a Gibbs sampling algorithm implemented in R. This procedure takes parameters: the response vector y , predictor matrix X , the number of MCMC iterations, the burn-in period, the precision parameter, each of which is initialized with the default values described earlier, with the possibility of modification. The output of this procedure includes estimates of the posterior means of all model unknowns and the mean of predictive distribution resulting from the MCMC algorithm. This procedure is included as supplementary material.

SIMULATION

Simulation study was conducted to evaluate the performance of the proposed method under skewness and contamination by outliers. We generated 100 observations from a Beta distribution. In addition to Beta-distributed responses, observations were generated from a logit-normal distribution to assess the robustness of the proposed method under distributional

misspecification. Since the BRQBR model assumes a Beta response, satisfactory performance on logit-normal data indicates that the method remains reliable when the true underlying distribution differs from the assumed Beta model. We considered that the number of covariates for both the mean and dispersion regressions were equal, taking $(p - 1) = (q - 1)$ with a total of 1000 dataset was simulated.

To ensure the presence of skewness in the response variable, we use the same values in Pérez-Rodríguez et al. (2022) as skewness factor = $\{0.75, 0.95, 0.9999\}$ as follows:

$$a = \frac{\mu\phi}{\text{skew} - \text{factor}}, \quad \beta = (1 - \mu)\phi \cdot \text{skew} - \text{factor}$$

Robustness was assessed by introducing contamination levels of 5% and 10% into the simulated datasets. Model performance was evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), and residual standard deviation (SD). A robust model is expected to exhibit smaller increases in these measures as the proportion of outliers increases.

Figure 1 illustrates the shapes of the simulated distributions under these skew-factors where Beta distribution has a positive skewness.

THE REAL DATASET

This study examines the factors affecting moisture content in a seaweed drying system using the V-Groove

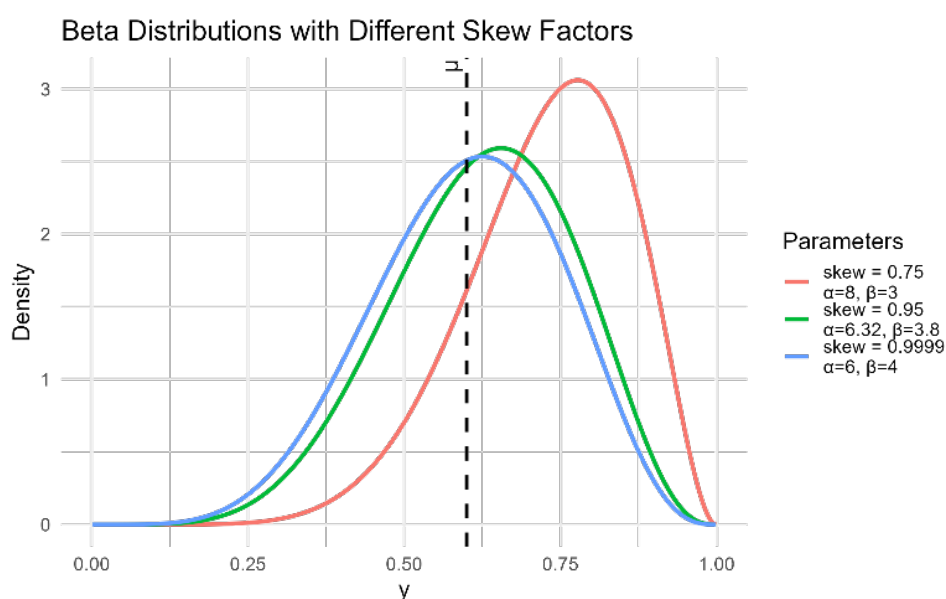


FIGURE 1. The positive skewness of Beta distribution with skew-factors

dataset, which includes hourly measurements of solar radiation, temperature, humidity, and moisture content. These data were collected from an experimental seaweed drying system and subsequently prepared for analysis through hierarchical clustering. In this work, we use a dataset of 1,924 observations to investigate the effects of twenty-nine independent variables, as listed in Table 1. This dataset was originally introduced by Mukhtar Eri et al. (2021). Figure 2 presents the variable-clustering dendrogram, while Table 2 summarizes the hierarchical clustering results, grouping the twenty-nine variables into six clusters and reporting the eigenvalues and variance explained by each cluster's principal component. Descriptive statistics are provided in Table 3, which show that the response variable Y is positively skewed, indicating a long right tail.

CROSS-VALIDATION ANALYSIS

We fitted BRQBR model and Beta regression estimated via MLE and selected three quantiles: $\theta = \{0.25, 0.50, 0.75\}$. We performed cross-validation analysis to evaluate the predictive ability of the proposed models. For each environment we generated 10-folds with 90% of the observations in training, and 10% in the testing 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 (Nascimento et al. 2017). Here we propose using the validation set to tune this parameter; this approach is widely used in machine learning (Pérez-Rodríguez et al. 2022). BRQBR model was fitted with 90% of observations in the training set, and the 10% of observations in the validation set were used to compute the mean squared error and mean absolute error, which can be used as a criterion for selecting models in cross-validation settings for each distribution. MLE-based Beta regression were fitted using 90% of observations in the training set and 10% of observations in the validation set for both Beta and logit-normal distributions.

RESULTS AND DISCUSSION

SIMULATION RESULTS

Table 4 presents simulation results under different levels of skewness and contamination. The values of MSE and MAE in Table 4 showed that the proposed BRQBR model consistently outperformed the MLE-based Beta regression approach in the two distributions where there are substantial differences between both. The MSE and MAE values obtained by BRQBR were substantially smaller than those obtained by the MLE approach. Figure 3 also shows the differences between the values of MSE and MAE in the two models in each

distribution. BRQBR also maintained superior performance in the presence of outliers. Table 5 shows the values of standard deviation in each quantile of both models with the skewness issue and outlier. BRQBR was significantly superior again compared with MLE-based Beta regression where most of the values in BRQBR are smaller than those in MLE in Beta and logit-normal distribution. Figure 4 indicates the difference between the standard deviation SD for the residuals for each quantile and the proposed method is performed the MLE in each distribution.

REAL DATASET

Table 6 and Figure 5 present the results of the mean squared error and mean absolute error for the validation set for the BRQBR and MLE-based Beta regression for Clustering's. v-Groove data. The seaweed drying dataset exhibits strong right skewness in the moisture content variable. BRQBR consistently produced smaller MSE and MAE than MLE for $\tau = 0.50$ and $\tau = 0.75$, indicating superior performance in the center and upper tail of the moisture distribution. This is expected because skewed responses with long right tails benefit from quantile-specific modeling and shrinkage priors that stabilize the estimates.

At $\tau = 0.25$, the MLE-based Beta regression estimator performed slightly better, likely because lower quantiles are less affected by the heavy right tail, and the estimated values via MLE is less sensitive when contamination is near the upper range rather than the lower quantile. Overall, BRQBR demonstrated greater robustness and stability, particularly in regions of the conditional distribution most influenced by skewness and outliers.

DISCUSSION

We introduced a robust Bayesian regularized quantile Beta regression model for analyzing skewed bounded data in the presence of outliers. Simulation results indicate that the proposed method provides more accurate parameter estimation and prediction. The advantage of the model is more evident when the data is more skewed, and it works well even in the presence of outliers; equivalent results were reported by Gianola et al. (2018), Montesinos-López et al. (2019) and Pérez-Rodríguez et al. (2022) but with Asymmetric Laplace Distribution. We simulated skewed and bounded data in (0,1) in two distributions to prove the robustness of our Bayesian regularized Beta regression where the results of MSE and MAE of our model are more accurate in most skew-factors values even if there are outliers or not. The same comparison was proposed in Bourguignon, Gallardo and Saulo (2024), where they proposed the model of quantile Beta regression with GB3 distribution but without addressing the skewness issue.

TABLE 1. The definition of the variables used in the dataset

Variables	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) Infront of (Inside Chamber)
T20, T23, T24, T25, T28	Temperature (°C) from solar collector to chamber

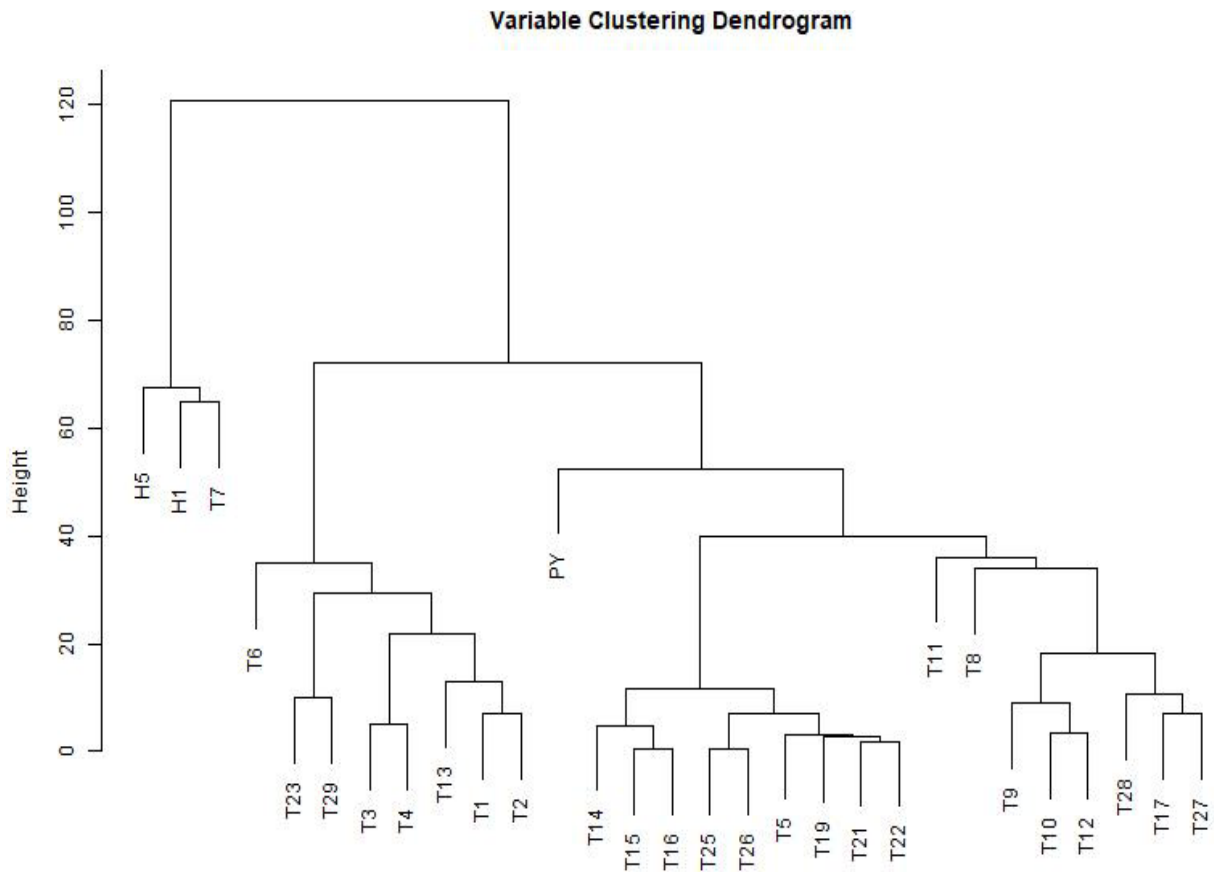


FIGURE 2. The variable clustering dendrogram

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

Clusters	No. of variables	Eigen value	percentage of variance	Cumulative of percentage of 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	
Cluster5	17	1.578449e+01	92.84994	92.84994	T5 T8 T14 T15 T16 T21 T22 T9 T10 T11 T12 T17 T19 T25 T26 T27 T28
		5.436771e-01	3.198101	96.04804	
		3.223073e-01	1.895925	97.94397	
		2.491098e-01	1.465352	99.40932	
		4.169977e-02	0.2452927	99.65461	
		2.389699e-02	0.1405705	99.79518	
		1.902288e-02	0.1118993	99.90708	
		5.773038e-03	0.03395905	99.94104	
		5.392559e-03	0.03172094	99.97276	
		1.661366e-03	9.772743e-03	99.98253	
		1.147210e-03	6.748295e-03	99.98928	
		7.876760e-04	4.633388e-03	99.99392	
		5.258986e-04	3.093521e-03	99.99701	
		3.690737e-04	2.171022e-03	99.99918	
		9.575735e-05	5.632785e-04	99.99974	
		3.613134e-05	2.125373e-04	99.99996	
		7.398697e-06	4.352175e-05	100	
Cluster6	1	-	-	-	T7

TABLE 3. Descriptive statistics

Variables	Mean	SD	Skewness	Kurtosis
Y	0.55234	0.1654	0.848239	-0.87574
H1	92.74993	170.1426	2.777199	7.465287
H5	34.7768	17.19086	1.920551	3.823985
PY	330.3889	255.7499	0.564006	-0.78236
PC1-Cluster4	2.7E-12	2.69398	-0.22921	-0.4899
PC1-Cluster5	1.41E-11	3.97297	-0.11656	-0.79015
T7	40.71594	5.22609	1.976757	8.244915

TABLE 4. Mean square error and mean absolute error for BRQBR and MLE for each quantile in condition with skew factor and outlier for Beta and Logit-normal distributions (The smallest value in Bold face)

Distribution	Skew factor	outlier	MSE-BRQBR			MSE-MLE		
	τ		0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.017195	0.017459	0.015944	0.011243	0.039974	0.085413
		5%	0.015691	0.0183	0.018992	0.004984	0.03356	0.098527
		10%	0.015816	0.01853	0.019279	0.007224	0.029553	0.101581
	0.95	0%	0.001416	0.001497	0.001419	0.026507	0.012865	0.040714
		5%	0.002475	0.002783	0.002844	0.027444	0.009881	0.045385
		10%	0.004856	0.005423	0.005553	0.03913	0.002857	0.049809
	0.999	0%	0.001314	0.001393	0.001334	0.034217	0.014068	0.033643
		5%	0.004597	0.004843	0.00462	0.043986	0.01276	0.039303
		10%	0.003807	0.004196	0.00425	0.031986	0.006006	0.049715
Distribution	Skew factor	outlier	MSE-BRQBR			MSE-MLE		
	τ		0.25	0.50	0.75	0.25	0.50	0.75
Logit normal	0.75	0%	0.001499	0.001142	0.001234	0.042158	0.010657	0.037142
		5%	0.006721	0.005293	0.004536	0.061086	0.007926	0.038666
		10%	0.004342	0.004454	0.00549	0.043691	0.003473	0.058872
	0.95	0%	0.000771	0.000378	0.000419	0.042773	0.010495	0.036563
		5%	0.001321	0.001305	0.001684	0.037242	0.01014	0.043172
		10%	0.004624	0.005037	0.006244	0.041583	0.007441	0.058811
	0.999	0%	0.001504	0.00115	0.001096	0.042379	0.012235	0.035384
		5%	0.003325	0.002413	0.002331	0.052332	0.005983	0.039534
		10%	0.003795	0.003185	0.003348	0.04932	0.00601	0.045886
Distribution	Skew factor	outlier	MAE-BRQBR			MAE-MLE		
	τ		0.25	0.50	0.75	0.25	0.50	0.75
Beta	0.75	0%	0.124082	0.124538	0.117381	0.092522	0.175902	0.268602
		5%	0.109959	0.116258	0.115313	0.059973	0.170666	0.290794
		10%	0.109351	0.11586	0.115124	0.074036	0.159684	0.295433
	0.95	0%	0.029817	0.030672	0.029621	0.146613	0.10156	0.18557
		5%	0.040391	0.042425	0.041896	0.148967	0.086185	0.194356
		10%	0.060693	0.06359	0.062632	0.186561	0.045622	0.211459
	0.999	0%	0.028739	0.029843	0.029114	0.166199	0.104527	0.165981
		5%	0.054334	0.05622	0.054667	0.187494	0.095936	0.176217
		10%	0.052736	0.055166	0.054306	0.164726	0.066484	0.206112
Distribution	Skew factor	outlier	MAE-BRQBR			MAE-MLE		
	τ		0.25	0.50	0.75	0.25	0.50	0.75
Logit normal	0.75	0%	0.03157	0.025982	0.027351	0.190976	0.090643	0.178004
		5%	0.063787	0.060189	0.056099	0.234013	0.075708	0.183384
		10%	0.055622	0.058382	0.059736	0.199927	0.050907	0.231428
	0.95	0%	0.023483	0.015392	0.016698	0.19363	0.090518	0.177646
		5%	0.029572	0.028441	0.032602	0.179203	0.089393	0.192836
		10%	0.055841	0.058726	0.061074	0.190399	0.070902	0.227861
	0.999	0%	0.031445	0.026141	0.026432	0.192034	0.099652	0.173361
		5%	0.045327	0.041825	0.039051	0.218227	0.066713	0.189538
		10%	0.047579	0.046371	0.045889	0.211393	0.066038	0.203747

Results obtained from the real dataset were consistent with those observed in the simulation study, where in cross-validation, the BRQBR model was better in the dataset. The proposed model is a robust alternative to the maximum likelihood estimator MLE, and its predictive ability is at the same level when no outliers are present. We considered only three quantiles, $\theta = \{0.25, 0.50, 0.75\}$; a similar approach was employed by Nascimento et al. (2017) and Pérez-

Rodríguez et al. (2022). For the results of the real dataset in the seaweed drying case, we proved that the proposed model handles skewness, kurtosis, and outliers much better, consistently producing smaller prediction errors. We also developed an algorithmic approach that allows selecting the quantile to be used to predict individuals in the testing set; this approach is based on cross-validation and is widely used in machine learning (Bishop 2006).

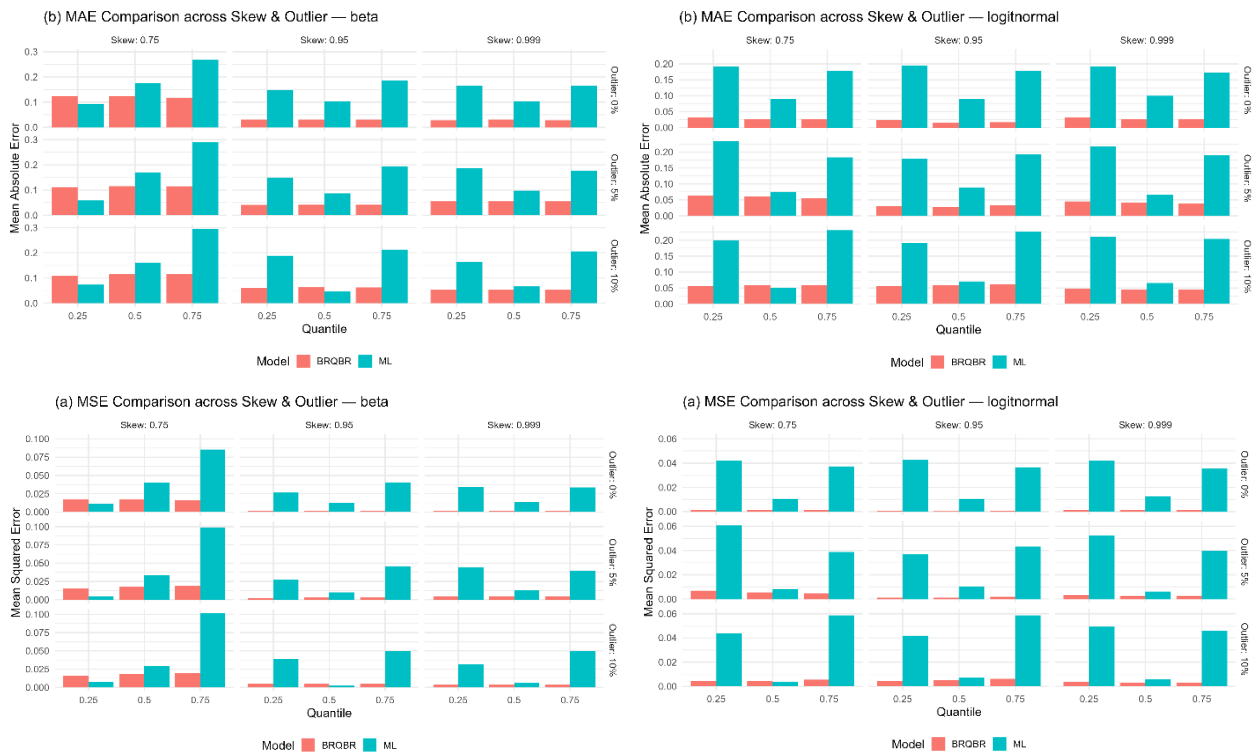


FIGURE 3. Differences in MSE and MAE across quantiles and contamination levels

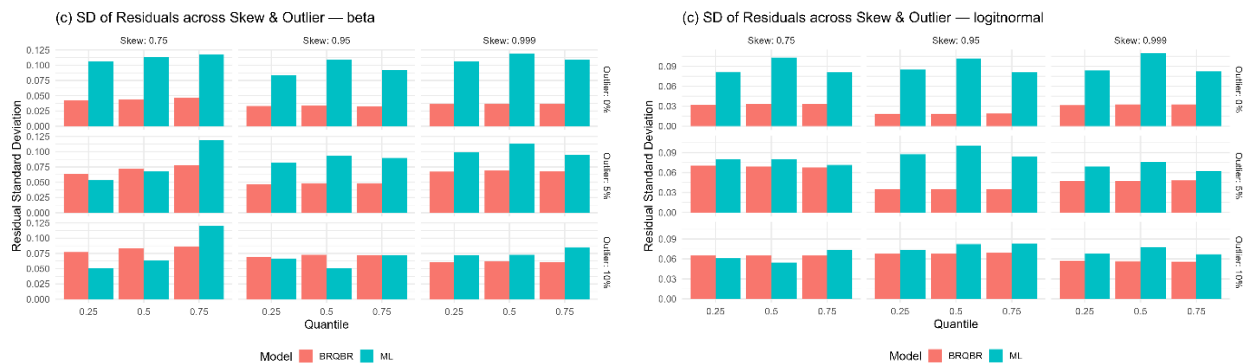


FIGURE 4. Differences in residual standard deviations across quantiles and contamination levels

TABLE 5. Standard deviation of residuals for BRQBR and MLE for each quantile in condition with skew factor and outlier for Beta and Logit-normal distributions (The smallest value in Bold face)

Distribution	Skew factor	outlier	SD-BRQBR			SD-MLE		
Beta	τ		0.25	0.50	0.75	0.25	0.50	0.75
		0%	0.042518	0.04426	0.046654	0.106298	0.112874	0.117589
	0.75	5%	0.063948	0.072234	0.077578	0.053127	0.067679	0.118477
		10%	0.076947	0.083399	0.086135	0.050588	0.063948	0.119886
		0%	0.033127	0.033866	0.032866	0.083673	0.109468	0.091869
	0.95	5%	0.046239	0.048015	0.047615	0.082487	0.09355	0.089886
		10%	0.069687	0.072778	0.072251	0.066119	0.050456	0.071553
		0%	0.036341	0.037382	0.036527	0.106072	0.118906	0.10965
	0.999	5%	0.067403	0.069565	0.068126	0.098996	0.11291	0.095014
		10%	0.060236	0.061699	0.060355	0.071803	0.0728	0.085318
Logit normal	τ		0.25	0.50	0.75	0.25	0.50	0.75
		0%	0.068126	0.032449	0.033337	0.095014	0.081616	0.10322
	0.75	5%	0.065509	0.070269	0.068953	0.073075	0.079724	0.080478
		10%	0.03383	0.065763	0.06522	0.081314	0.061147	0.054836
		0%	0.017889	0.018159	0.019061	0.085651	0.102166	0.081041
	0.95	5%	0.035606	0.035429	0.034743	0.087665	0.100671	0.084403
		10%	0.068144	0.068601	0.069234	0.07369	0.082508	0.083218
		0%	0.031387	0.032731	0.032799	0.083886	0.110034	0.082949
	0.999	5%	0.047642	0.047113	0.04811	0.068797	0.075917	0.06209
		10%	0.057246	0.056409	0.056123	0.068237	0.077547	0.067337

TABLE 6. Mean squared error, mean absolute error and standard deviation for the testing dataset from cross-validation analysis (The smallest value in Bold face)

Method	Quantiles	MSE	MAE	COR ($y, \hat{y}$)
BRQBR	0.25	0.017455	0.107202	0.745329
	0.50	0.012385	0.088993	0.741792
	0.75	0.019641	0.115832	0.737642
MLE	0.25	0.10423	0.302834	0.758082
	0.50	0.022533	0.119436	0.734462
	0.75	0.11965	0.322875	0.661396

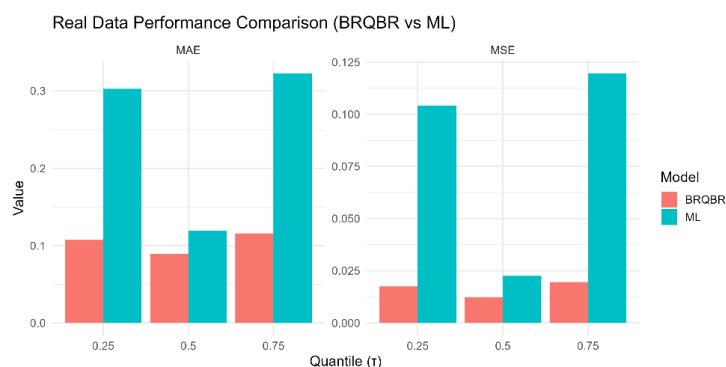


FIGURE 5. Comparison of MSE and MAE for the real dataset

CONCLUSIONS

This paper developed the Bayesian regularized quantile Beta regression (BRQBR). The proposed method overcomes limitations of existing approaches by integrating quantile Beta regression with Bayesian regularization. This method addresses the problem of skewness in the data and improves flexibility in modeling complex and over-dispersed data. Bayesian inference was used to achieve high accuracy using prior distributions and proposed initial values, imposing variable variances for the main regression parameters.

Monte Carlo simulations showed that the proposed method outperformed the MLE-based Beta regression in various situations, as evidenced by the decrease in the mean error of the mean coefficient vector. Three levels of skewness and two contamination rates were considered across two distributions. The proposed model outperformed in most cases of skewness even if there are outliers or not.

A real seaweed drying dataset collected in Malaysia was used to evaluate the practical performance of the proposed method. The response data are positively skewed. BRQBR yields smaller MSE and MAE compared to MLE-based Beta regression across all quantiles. The improvement of the proposed model is also shown through the Pearson correlation. Additionally, MLE method sometimes produces unstable fits, while BRQBR stabilizes predictions.

Future research may explore expanding the scope of this work by comparing the model with other regularization methods such as Ridge or Lasso, as well as using the proposed model in various medical, financial, and economic studies, or studying the effects of extreme values on the estimated parameters.

ACKNOWLEDGMENTS

We would like to express our gratitude to the School of Mathematical Sciences, Universiti Sains Malaysia for providing the facilities for this study.

REFERENCES

- Abonazel, M.R., Algarnal, Z.Y., Awwad, F.A. & Taha, I.M. 2022. A new two-parameter estimator for Beta regression model: Method, simulation, and application. *Frontiers in Applied Mathematics and Statistics* 7: 780322. <https://doi.org/10.3389/fams.2021.780322>
- Akram, M.N., Amin, M., Elhassanein, A. & Ullah, M.A. 2022. A new modified ridge-type estimator for the beta regression model: Simulation and application. *AIMS Math.* 7(1): 1035-1057. doi:10.3934/math.2022062
- Bishop, C.M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin. Heidelberg. Germany: Springer-Verlag.
- Bourguignon, M., Gallardo, D.I. & Saulo, H. 2024. Parametric quantile beta regression model. *International Statistical Review* 92(1): 106-129. <https://doi.org/10.1111/insr.12564>
- Carvalho, C.M., Polson, N.G. & Scott, J.G. 2010. The horseshoe estimator for sparse signals. *Biometrika* 97(2): 465-480. <https://doi.org/10.1093/biomet/asq017>
- Espinheira, P.L., Ferrari, S.L. & Cribari-Neto, F. 2008. On Beta regression residuals. *Journal of Applied Statistics* 35(4): 407-419. <https://doi.org/10.1080/02664760701834931>
- Ferrari, S. & Cribari-Neto, F. 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31(7): 799-815. <https://doi.org/10.1080/0266476042000214501>
- Ferrari, S.L., Espinheira, P.L. & Cribari-Neto, F. 2011. Diagnostic tools in Beta regression with varying dispersion. *Statistica Neerlandica* 65(3): 337-351. <https://doi.org/10.1111/j.1467-9574.2011.00488.x>
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. 2013. *Bayesian Data Analysis*. 3rd ed. Boca Raton: CRC Press. <https://doi.org/10.1201/b16018>

- Gianola, D., Cecchinato, A., Naya, H. & Schön, C.C. 2018. Prediction of complex traits: Robust alternatives to best linear unbiased prediction. *Front. Genet.* 9: 195. <https://doi.org/10.3389/fgene.2018.00195>
- Karlsson, P., Månsson, K. & Kibria, B.G. 2020. A liu estimator for the beta regression model and its application to chemical data. *Journal of Chemometrics* 34(10): e3300. <https://doi.org/10.1002/cem.3300>
- Koenker, R. & Bassett, G. 1978. Regression quantiles. *Econometrica* 46(1): 33-50. <https://doi.org/10.2307/1913643>
- Li, Q., Xi, R. & Lin, N. 2010. Bayesian regularized quantile regression. *Bayesian Analysis* 5(3): 533-556. <https://doi.org/10.1214/10-ba52>
- McElreath, R. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2nd ed. Boca Raton: CRC Press. <https://doi.org/10.1201/9780429029608>
- Montesinos-López, A., Montesinos-López, O.A., Villa-Diharce, E.R., Gianola, D. & Crossa, J. 2019. A robust Bayesian genome-based median regression model. *Theoretical and Applied Genetics* 132: 1587-1606. <https://doi.org/10.1007/s00122-019-03303-6>
- Mukhtar Eri Suhaeri, Alimudin, Anam Javaid, Mohd. Tahir Ismail & Majid Khan Majahar Ali. 2021. Evaluation of clustering approach with Euclidean and Manhattan distance for outlier detection. *AIP Conference Proceedings* 2423: 070025. <https://doi.org/10.1063/5.0075570>
- Nascimento, M., e Silva, F.F., de Resende, M.D.V., Cruz, C.D., Nascimento, A.C.C., Viana, J.M.S., Azevedo, C.F. & Barroso, L.M.A. 2017. Regularized quantile regression applied to genome-enabled prediction of quantitative traits. *Genet. Mol. Res.* 16(1): 9538. doi: 10.4238/gmr16019538
- Pérez-Rodríguez, P., Montesinos-López, O.A., Montesinos-López, A. & Crossa, J. 2022. Bayesian regularized quantile regression: A robust alternative for genome-based prediction of skewed data. *The Crop Journal* 8(1): 713-722. <https://doi.org/10.1016/j.cj.2020.04.009>
- Piironen, J. & Vehtari, A. 2017. On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. PMLR 54: 905-913.
- Simas, A.B., Barreto-Souza, W. & Rocha, A.V. 2010. Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis* 54(2): 348-366. <https://doi.org/10.1016/j.csda.2009.08.017>
- Smithson, M. & Verkuilen, J. 2006. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11(1): 54-71. <https://doi.org/10.1037/1082-989X.11.1.54>
- Tsionas, E.G. 2003. Bayesian quantile inference. *Journal of Statistical Computation and Simulation* 73(9): 659-674. <https://doi.org/10.1080/0094965031000064463>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.C. 2021. Rank- normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis* 16(2): 667-718. <https://doi.org/10.1214/20-BA1221>
- Yu, K. & Moyeed, R.A. 2001. Bayesian quantile regression. *Statistics and Probability Letters* 54(4): 437-447. [https://doi.org/10.1016/S0167-7152\(01\)00124-9](https://doi.org/10.1016/S0167-7152(01)00124-9)

*Corresponding author; email: majidkhanmajaharali@usm.my