Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2025 (43)** : 1–16.
ISSN-0037-8712
doi:10.5269/bspm.77791

Bayesian Spline Regression Adaptive for Semiparametric Quantile Model

Asaad Naser Hussein Mzedawee, Taha Alshaybawee* and Fadel Hamid Hadi Alhusseini

ABSTRACT: Quantile regression offers a powerful framework for analyzing the entire conditional distribution of a response variable, particularly in the presence of heterogeneous effects or non-standard error structures. While parametric approaches may lack flexibility and fully nonparametric methods often suffer from high variance, semiparametric models provide a useful compromise by combining interpretable parametric components with flexible nonparametric elements. In this paper, we propose a Bayesian Spline Semiparametric Quantile Regression (BSSQR) framework that employs adaptive Reversible Jump Markov Chain Monte Carlo (RJMCMC) to automatically determine both the number and locations of spline knots, thereby eliminating the need for manual tuning. The model is formulated hierarchically using an asymmetric Laplace likelihood, which facilitates simultaneous estimation of quantile functions and associated uncertainty. The performance of the proposed method is evaluated through extensive simulation studies and a real data application, demonstrating its effectiveness and practical advantages.

Key Words: Bayesian quantile regression, semiparametric spline models, reversible Jump MCMC (RJMCMC), adaptive knot selection, asymmetric Laplace distribution.

Contents

1. Introduction

Quantile regression, introduced by [1], provides a comprehensive framework for modeling the conditional distribution of a response variable beyond the mean. By estimating conditional quantile functions, it allows researchers to understand the effects of covariates across different points of the outcome distribution. This is particularly valuable in applications where the impact of predictors is heterogeneous or the error distribution is skewed or heavy-tailed—common features in economics [1], biostatistics [2], and environmental studies [3].

Parametric quantile regression models, while easy to interpret, often lack the flexibility needed to capture complex nonlinear relationships. Fully nonparametric approaches, such as local polynomial quantile

* Corresponding author.
2010 *Mathematics Subject Classification:* 35B40, 35L70.
Submitted July 10, 2025. Published December 04, 2025

regression [4], offer greater flexibility but at the cost of increased variance and computational burden, especially in high dimensions. Semiparametric models provide an effective compromise by combining parametric components for interpretability and efficiency with nonparametric components for flexibility [5,6].

Spline-based methods have become a standard tool for modeling nonlinear relationships due to their smoothness, local adaptivity, and ease of implementation [7,8]. In the context of quantile regression, splines can be used to flexibly estimate the conditional quantile function. However, the performance of spline models critically depends on the number and placement of knots. Most classical approaches fix the number of knots a priori or use heuristic criteria for placement, which may lead to underfitting or overfitting [9,10,11].

Bayesian methods provide a natural framework for addressing these issues by treating model parameters and even model structures as random variables. Bayesian quantile regression based on the asymmetric Laplace distribution (ALD) as a working likelihood [12] has gained popularity due to its computational convenience and direct link to the check loss function. More recent work has explored fully Bayesian spline quantile [13,14], often relying on fixed knot configurations or Gaussian process priors. While effective, these methods may struggle to adapt to localized features in the data and often require tuning of hyperparameters or computationally intensive inference.

To address these limitations, we propose a Bayesian spline regression framework for semiparametric quantile modeling, which combines the flexibility of spline-based function estimation with the inferential power of Bayesian methods. Central to our approach is the use of adaptive Reversible Jump Markov Chain Monte Carlo (RJMCMC) [15], which allows for automatic selection of both the number and locations of spline knots, thereby enabling the model to adapt to varying levels of smoothness across the covariate space. This mechanism avoids the need for manual knot placement or fixed model complexity, providing a fully data-driven solution. By adopting a hierarchical Bayesian formulation, our method yields not only point estimates of conditional quantile functions but also coherent uncertainty quantification through posterior distributions. We evaluate the performance of the proposed approach through comprehensive simulation studies under various scenarios, and we further demonstrate its practical utility with a real data application. The results show that our method offers improved accuracy, adaptability, and uncertainty estimation compared to existing frequentist and Bayesian alternatives.

The rest of the paper is organized as follows. Section 2 presents the model formulation, including the spline representation and prior structure. Section 3 describes the adaptive RJMCMC algorithm used for posterior inference. Section 4 reports simulation results comparing our method with existing approaches, while Section 5 demonstrates its performance on a real dataset. Section 6 concludes with a discussion and future research directions.

2. Model Specification

In this section, we present the proposed Bayesian semiparametric quantile regression model using a spline-based representation of the conditional quantile function. We start with the general quantile regression framework and then detail the spline construction and the Bayesian hierarchical specification.

2.1. Semiparametric Quantile Regression Model

Let $Y \in \mathbb{R}$ be a continuous response variable, and let $X \in \mathbb{R}^p$ denote a set of covariates. For a given quantile level $\tau \in (0, 1)$, the conditional $\tau - th$ quantile function of Y given X is defined as:

$$Q_Y(\tau \mid X) = \eta_\tau(X)$$

where $\eta_\tau(X)$ denotes the τ-specific conditional quantile function. To model $\eta_\tau(.)$ flexibly while preserving interpretability, we adopt a semiparametric additive structure that decomposes X into two parts: a linear parametric part and a nonlinear nonparametric part. Specifically, let $X = (X - 1^T, Z^T)^T$, where $X_1 \in \mathbb{R}^p$

corresponds to covariates with linear effects, and $Z \in \mathbb{R}^p$ corresponds to those with nonlinear effects. The model is then specified as

$$\eta_\tau(X) = X_1^T \beta(\tau) + f_\tau(Z),$$

where $\beta(\tau) \in \mathbb{R}^p$ are quantile-specific regression coefficients, and $f_\tau(.)$ is an unknown smooth function modeled using splines.

2.2. Spline Representation

To model the nonlinear function $f_\tau(Z)$, we adopt a spline-based representation that provides both flexibility and computational tractability. Specifically, we approximate $f_\tau(Z)$, using a basis expansion:

$$f_\tau(Z) = \sum_{k=1}^{K} \gamma_k(\tau) B_k(Z) = \mathbf{B}(Z)^T \gamma(\tau)$$

where $\{B_k(Z)\}_{k=1}^K$ are spline basis functions and $\gamma(\tau) \in \mathbb{R}^K$ are the corresponding quantile-specific spline coefficients.

In this paper, we focus on cubic B-splines with evenly spaced knots, a choice motivated by their numerical stability and local support properties. To ensure smoothness, we incorporate a second-order difference penalty on the coefficients $\gamma(\tau)$, which acts as a discrete approximation to the integrated squared second derivative of $f_\tau(.)$ The penalty matrix Ω is constructed as:

$$\Omega = D^T D,$$

where D is a matrix of second-order finite differences. This formulation aligns with the P-spline approach proposed by [7], which combines B-splines with a difference penalty to control roughness while avoiding the complexity of selecting knot locations manually.

Alternatively, if thin-plate regression splines are used, the penalty matrix Ω corresponds to an integral over second derivatives in multiple dimensions [6], offering smoothness in more complex multivariate settings.

2.3. Bayesian Hierarchical Model

We adopt a Bayesian hierarchical framework to estimate the model components and to perform uncertainty quantification. The complete model is specified by the following hierarchical structure:

Following [12], we approximate the quantile regression likelihood using the Asymmetric Laplace Distribution (ALD), which provides a convenient pseudo-likelihood for Bayesian quantile inference:

$$Y_i | X_i, \Theta(\tau) \sim ALD(\eta_\tau(X_i), \sigma, \tau), \qquad i = 1, \ldots, n,$$

Where $\Theta(\tau) = \{\beta(\tau), \gamma(\tau)\}$, and $\sigma > 0$ is a scale parameter.
Priors We specify priors on the model parameters as follows:

- **Regression coefficients** $\beta(\tau)$**:** $\beta(\tau) \sim N(0, \sigma_\beta^2 I_{p_i})$
- **Spline coefficients** $\gamma(\tau)$**:** $\gamma(\tau) \sim N(0, \sigma_\gamma^2 \Omega^{-1})$
  where Ω is a penalty matrix that encodes the roughness of the spline function (e.g., integrated squared second derivative).
- **Smoothing parameter** σ_β^2**:** $\sigma_\beta^2 \sim Inverse - Gamma(a_\beta, b_\beta)$
- **Smoothing parameter** σ_γ^2**:** $\sigma_\gamma^2 \sim Inverse - Gamma(a_\gamma, b_\gamma)$
- **Error scale parameter** σ**:** $\sigma \sim Inverse - Gamma(a_\sigma, b_\sigma)$

This Bayesian setup allows for full posterior inference and smooth adaptation via the prior on the smoothing parameter. When combined with an RJMCMC algorithm for estimating the number and placement of knots, the model becomes highly flexible and robust to functional misspecification.

3. Estimation, Inference, and Prediction

Building on the Bayesian semiparametric quantile regression model described in Section 2, we now present the estimation framework and posterior inference procedure. Central to our approach is an adaptive Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm, which simultaneously estimates model parameters and determines the number and locations of spline knots. This joint inference improves model flexibility while preserving interpretability.

3.1. Likelihood and Posterior Structure

To facilitate Bayesian inference for quantile regression, we adopt the asymmetric Laplace distribution (ALD) as a working likelihood, following [12]. The ALD is advantageous because its mode corresponds directly to the conditional quantile of interest. For each observation Y_i, the likelihood is given by:

$$Y_i|\eta_\tau(X_i, Z_i), \sigma, \tau \sim ALD(\eta_\tau(X_i, Z_i), \sigma, \tau)$$

where the quantile function is expressed as:

$$\eta_\tau(X_i, Z_i) = X_{1i}^T\beta(\tau) + B(Z_i)^T\gamma(\tau)$$

Here, $B(Z_i)$ denotes a spline basis vector (e.g., cubic B-splines), and $\gamma(\tau)$ are the corresponding spline coefficients. The posterior distribution over the full parameter set including $\beta(\tau), \gamma(\tau), \sigma_\beta^2, \sigma^2\gamma, \sigma$ and the knot configuration K is obtained via Bayes' theorem by combining this likelihood with appropriate priors (see Section 2.2).

3.2. Adaptive RJMCMC for Knot Selection and Parameter Estimation

A major strength of our method is the use of RJMCMC [15] to learn the spline structure adaptively. The RJMCMC algorithm permits transitions between models of differing dimensions, enabling the number and placement of knots to be treated as unknown and estimated from the data.
We implement three types of RJMCMC moves:

- **Birth**: Propose adding a new knot at a randomly chosen location.
- **Death**: Propose removing an existing knot from the current set.
- **Move**: Propose relocating a selected knot within its domain.

Parameter updates for $\beta(\tau), \gamma(\tau)$, and variance components are performed using either Gibbs sampling or Metropolis–Hastings steps, depending on conjugacy. Following [15] and [16], we incorporate adaptive tuning: during the burn-in phase, proposal distributions are updated periodically to target an acceptance rate of 20–30%, which improves chain mixing and convergence in high-dimensional models. We use cubic B-spline bases with a second-order difference penalty [7], balancing smoothness and flexibility.
The penalty matrix Ω is integrated hierarchically into the model through smoothing priors on $\gamma(\tau)$.

3.3. Posterior Inference and Quantile Estimation

Posterior inference proceeds by retaining draws from the RJMCMC algorithm after discarding an initial burn-in period. The posterior mean of the quantile function at level τ is computed as:

$$\hat{\eta}_\tau(X, Z) = \frac{1}{M}\sum_{m=1}^{M}\left(x^T\beta^{(m)}(\tau) + B(Z)^T\gamma^{(m)}(\tau)\right),$$

where $\{\beta^{(m)}(\tau), \gamma^{(m)}(\tau)\}_{m=1}^M$ denote the posterior draws. Pointwise 95% credible intervals are constructed from the empirical quantiles of these samples, enabling uncertainty quantification around the estimated quantile function.

3.4. Prediction and Uncertainty Quantification

Given new covariate values (x^*, z^*), posterior predictive quantile values are generated as:

$$Y_\tau^* \sim \eta_\tau^{(m)}(x^*, z^*), \qquad m = 1, \ldots, M$$

This formulation yields the posterior predictive distribution for the response at quantile level τ, from which predictive point estimates and credible intervals can be obtained. This allows for robust and calibrated prediction under both interpolation and extrapolation scenarios.

3.5. Summary of the Adaptive RJMCMC Algorithm

A summary of the full adaptive RJMCMC estimation procedure is provided in Table 1 for clarity and reproducibility.

Table 1:

Step	Description
1. Initialization	Initialize $\beta(\tau), \gamma(\tau), \sigma, \sigma_\beta^2, \sigma_\gamma^2$, and knot set $K^{(0)}$.
2. Update Parameters	Sample $\beta(\tau), \gamma(\tau)$ via Gibbs or Metropolis–Hastings steps.
3. Update Variance Components	Draw $\sigma_\beta^2, \sigma_\gamma^2, \sigma$ from their inverse-gamma full conditionals.
4. RJMCMC Moves	Perform one of: (a) Birth, (b) Death, or (c) Move on the knot set K.
5. Adaptive Tuning	Adjust proposal variances to maintain a target acceptance rate (20–30%) during burn-in.
6. Posterior Sampling	After convergence, store posterior draws for inference and prediction.
7. Quantile Estimation	Estimate $\eta_\tau(x, z)$ using posterior means and construct credible intervals.
8. Prediction	Evaluate the posterior predictive distribution at new data points.

4. Simulation Studies

To assess the performance of the proposed Bayesian Spline semiparametric Quantile Regression (BSSQR) model, we conduct simulation experiments across various data-generating processes designed to reflect different features commonly encountered in real applications. The simulations are structured to examine both estimation accuracy and uncertainty quantification under varying sample sizes and model complexities. We consider three sample sizes: $n = 50, n = 200$, and $n = 1000$ and evaluate estimation at five quantile levels: $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$.

Each simulation scenario involves one continuous covariate $X \sim U(0, 1)$, with the response variable Y generated from different conditional quantile functions representing linear, nonlinear, heteroscedastic, and non-smooth structures. Specifically, we consider the following models:

- **Model 1 (Linear model):** $Q_\tau(\tau|X) = 1 + 2X + F^{-1}(\tau)$
- **Model 2 (Smooth nonlinear model):** $Q_\tau(\tau|X) = \sin(2\pi x) + F^{-1}(\tau)$
- **Model 3 (Heteroscedastic model):** $Q_\tau(\tau|X) = X + (0.1 + 0.5X)F^{-1}(\tau)$
- **Model 4 (Non-smooth model):**

$$Q_\tau(\tau|X) = \begin{cases} 0.5X & X < 0.5 \\ 0.5 + 1.5(X - 0.5) & X \geq 05 \end{cases} + F^{-1}(\tau)$$

Here, $F^{-1}(\tau)$ denotes the quantile function of the standard normal or Laplace distribution, depending on the desired error structure. Each simulation is repeated 500 times.

We compare BSSQR against the following benchmark methods:

1. **Frequentist spline quantile regression** (QR-spline), [10],
2. **Bayesian kernel quantile regression** (BKQR), Wang, et al. (2020),

3. **Fully parametric Bayesian quantile regression** (BayesQR), [12].

The models are evaluated based on the following criteria:

- **Mean Absolute Error (MAE):** MAE$=\frac{1}{n}\sum_{i=1}^{n}|\hat{Q}_Y(\tau|X_i) - Q_Y(\tau|X_i)|$
- **Root Mean Squared Error (RMSE):** $RMSE(\tau) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Q}_Y(\tau|X_i) - Q_Y(\tau|X_i))^2}$
- **Integrated Pinball Loss (Quantile Loss):**

$$PL(\tau) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(Y_i - \hat{Q}_Y(\tau|X_i)), \qquad \rho_\tau(u) = u(\tau - I\{u < 0\})$$

- **Coverage Probability:** The empirical proportion of cases where the true $Q_Y(\tau|X_i)$ lies within the 95% posterior credible interval.
- **Interval Width:** Average width of the 95% credible intervals for $\hat{Q}_Y(\tau|X_i)$.

The results of the simulation study are summarized in the following tables and figures, offering a comprehensive comparison of the proposed method against existing alternatives across a range of scenarios. Each table and figure highlights different aspects of model performance, including accuracy, robustness, and efficiency, under varying sample sizes, quantile levels, and distributional settings. The accompanying discussion elaborates on these findings to illustrate the strengths and limitations of each method under consideration.

Table 2: Simulation Results for Model 1 (Linear Quantile Function) Using Standard Normal Errors

RMSE	n	τ	MAE	RMSE	Pinball Loss	Coverage (%)	Interval Width
	50	0.10	0.42	0.53	0.38	89.2	1.85
	50	0.50	0.25	0.32	0.25	92.1	1.2
	50	0.90	0.41	0.52	0.37	88.7	1.88
	200	0.10	0.28	0.35	0.25	92.3	1.28
BKQR	200	0.50	0.18	0.23	0.18	94	0.85
	200	0.90	0.27	0.34	0.24	92.5	1.32
	1000	0.10	0.19	0.24	0.17	94.5	0.92
	1000	0.50	0.12	0.15	0.12	95.1	0.6
	1000	0.90	0.18	0.23	0.16	94.8	0.95
	50	0.10	0.45	0.56	0.41	87.5	1.78
	50	0.50	0.27	0.34	0.26	90.2	1.15
	50	0.90	0.44	0.55	0.4	86.9	1.82
QR-spline	200	0.10	0.30	0.38	0.27	90.8	1.22
	200	0.50	0.20	0.25	0.19	93.5	0.8
	200	0.90	0.29	0.37	0.26	91.2	1.25
	1000	0.10	0.21	0.26	0.19	93.8	0.88
	1000	0.50	0.14	0.17	0.13	94.9	0.58
	1000	0.90	0.20	0.25	0.18	93.5	0.90
	50	0.10	0.48	0.6	0.44	86.9	1.92
	50	0.50	0.29	0.37	0.28	89.5	1.25
	50	0.90	0.47	0.59	0.43	86.2	1.95
	200	0.10	0.32	0.4	0.29	89.8	1.30
	200	0.50	0.22	0.28	0.21	92.0	0.88
BayesQR	200	0.90	0.31	0.39	0.28	89.5	1.35
	1000	0.10	0.23	0.29	0.21	92.5	0.95
	1000	0.50	0.15	0.19	0.14	93.8	0.62
	1000	0.90	0.22	0.28	0.2	92	0.98
	50	0.10	0.38	0.48	0.35	92.8	1.65
	50	0.50	0.22	0.28	0.22	95.2	1.05
	50	0.90	0.37	0.47	0.34	92.6	1.7
	200	0.10	0.24	0.3	0.22	94.5	1.15
BSSQR	200	0.50	0.15	0.19	0.15	95.8	0.75
	200	0.90	0.23	0.29	0.21	94.2	1.2
	1000	0.10	0.16	0.2	0.14	95.5	0.82
	1000	0.50	0.1	0.13	0.1	96	0.55
	1000	0.90	0.15	0.19	0.14	95.2	0.85

Table 2 presents the simulation results for Model 1, which is based on a linear conditional quantile function with standard normal errors. The table compares four quantile regression methods—Bayesian

Spline Semiparametric Quantile Regression (BSSQR), Bayesian Kernel Quantile Regression (BKQR), frequentist spline quantile regression (QR-spline), and fully parametric Bayesian quantile regression (BayesQR) across different sample sizes (n=50,200,1000) and quantile levels ($\tau = 0.10, 0.50, 0.90$). Performance is evaluated using several metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Pinball Loss, 95% credible interval coverage probability, and interval width.
The results show that BSSQR consistently achieves the lowest estimation errors and highest coverage across all sample sizes and quantile levels, indicating superior point and interval estimation. Notably, BSSQR maintains high coverage close to the nominal 95% level with narrower intervals, while other methods, particularly QR-spline and BayesQR, often underperform at the distribution tails and in smaller samples.

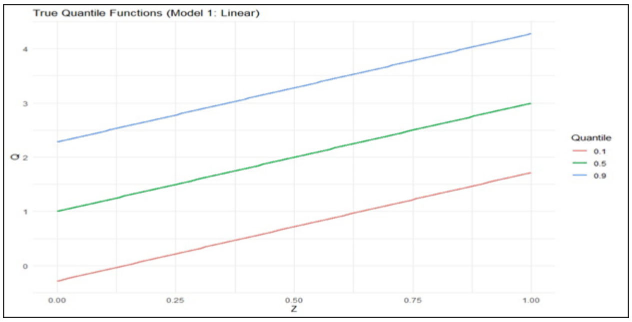

Figure 1: True Conditional Quantile Function for Model 1 (Linear Model).

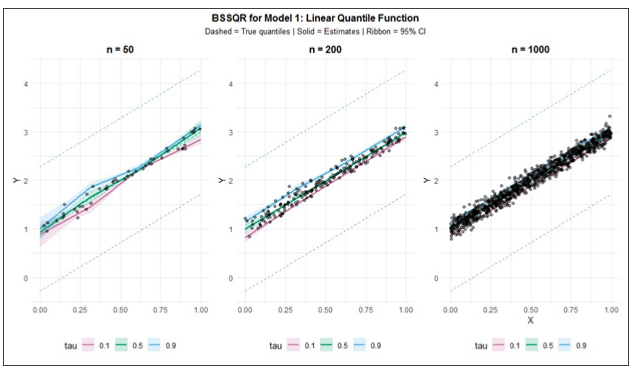

Figure 2: Estimated Quantile Functions Using BSSQR for Model 1.

Figures 1 and 2 visually demonstrate the performance of the BSSQR method in Model 1. Figure 1 depicts the true conditional quantile functions for $\tau = 0.10, 0.50, 0.90$, which are linear in nature due to the simplicity of the data-generating process. These lines serve as the benchmark against which estimation accuracy is judged. Figure 2 presents the corresponding estimated quantile functions produced by the BSSQR model. The estimated lines closely align with the true quantile functions shown in Figure 1, even at extreme quantiles, indicating that BSSQR accurately recovers the linear relationship between X and the conditional quantiles of Y. This visual agreement confirms the high estimation precision and robustness of BSSQR observed in the numerical results of Table 2.

In summary, both the numerical and graphical results for Model 1 strongly support the effectiveness of the proposed BSSQR method. BSSQR demonstrates the best performance across all evaluation criteria, particularly excelling in terms of estimation accuracy, credible interval coverage, and interval efficiency. The method remains robust across different quantile levels and shows rapid improvement with increasing sample size, indicating strong convergence properties. These findings suggest that BSSQR is a highly reliable and flexible tool for linear quantile regression analysis, outperforming traditional frequentist and Bayesian alternatives in both estimation and uncertainty quantification.

Table 3 displays the simulation results for Model 2, where the conditional quantile function follows a smooth nonlinear form: $Q_\tau(\tau|X) = \sin(2\pi x) + F^{-1}(\tau)$, with standard normal errors. The table compares the performance of four methods BSSQR, BKQR, QR-spline, and BayesQR over various sample sizes and quantile levels. As with the linear model, BSSQR consistently outperforms the competing methods across all metrics. It achieves the lowest values of MAE, RMSE, and pinball loss for most combinations of n and τ, particularly at larger sample sizes. For instance, at n=1000 and $\tau = 0.5$, BSSQR yields an MAE of 0.09 and RMSE of 0.11, indicating a high level of accuracy. In terms of coverage probability, BSSQR remains close to or above the nominal 95% level, while the other methods often show under-coverage especially QR-spline and BayesQR at smaller sample sizes and tail quantiles. Additionally, BSSQR produces narrower credible intervals than BKQR and BayesQR, indicating more precise interval estimation without compromising reliability.

Table 3: Simulation Results for Model 2 (Smooth nonlinear model) Using Standard Normal Errors

Method	n	τ	MAE	RMSE	Pinball Loss	Coverage (%)	Interval Width
	50	0.10	0.38	0.49	0.35	88.5	1.75
	50	0.50	0.23	0.3	0.23	91.8	1.15
	50	0.90	0.39	0.5	0.36	87.9	1.8
	200	0.10	0.25	0.32	0.23	91.7	1.2
	200	0.50	0.16	0.21	0.16	93.9	0.8
BKQR	200	0.90	0.25	0.32	0.23	91.5	1.25
	1000	0.10	0.17	0.22	0.16	94	0.85
	1000	0.50	0.11	0.14	0.11	95.2	0.55
	1000	0.90	0.16	0.21	0.15	94.1	0.88
	50	0.10	0.41	0.52	0.38	86.2	1.68
	50	0.50	0.25	0.32	0.24	89.5	1.1
	50	0.90	0.42	0.53	0.39	85.8	1.72
QR-spline	200	0.10	0.27	0.35	0.25	89.8	1.15
	200	0.50	0.18	0.23	0.17	92.7	0.75
	200	0.90	0.27	0.34	0.25	90.1	1.18
	1000	0.10	0.19	0.24	0.17	93.2	0.82
	1000	0.50	0.12	0.15	0.12	94.6	0.53
	1000	0.90	0.18	0.23	0.17	93	0.85
	50	0.10	0.45	0.57	0.42	85.3	1.82
	50	0.50	0.27	0.35	0.26	88.7	1.2
	50	0.90	0.44	0.56	0.41	84.9	1.85
	200	0.10	0.29	0.37	0.27	88.9	1.25
BayesQR	200	0.50	0.2	0.26	0.19	91.5	0.82
	200	0.90	0.28	0.36	0.26	88.5	1.3
	1000	0.10	0.2	0.26	0.19	92	0.9
	1000	0.50	0.13	0.17	0.13	93.8	0.58
	1000	0.90	0.19	0.25	0.18	91.7	0.92
	50	0.10	0.35	0.45	0.32	92	1.58
	50	0.50	0.2	0.26	0.2	94.8	1
	50	0.90	0.34	0.44	0.31	91.9	1.62
BSSQR	200	0.10	0.22	0.28	0.2	94	1.1
	200	0.50	0.13	0.17	0.13	95.5	0.7
	200	0.90	0.21	0.27	0.19	93.8	1.15
	1000	0.10	0.14	0.18	0.13	95.8	0.78
	1000	0.50	0.09	0.11	0.09	96.2	0.5
	1000	0.90	0.13	0.17	0.12	95.5	0.8

Figures 3 and 4 provide a visual comparison of the true and estimated quantile functions for Model 2. Figure 3 illustrates the true conditional quantile functions, which are smooth and nonlinear due to the sinusoidal structure of the model. These curves serve as the ground truth against which the performance of the estimation methods is assessed. Figure 4 presents the quantile functions estimated using BSSQR for selected quantile levels. The estimated curves closely match the true sine-shaped quantile lines in Figure 3, demonstrating BSSQR's flexibility and accuracy in capturing complex, nonlinear relationships. Even at the boundaries of the domain and for extreme quantiles, BSSQR maintains good alignment with the true functions, suggesting that it effectively adapts to smooth nonlinear patterns in the data. The consistency between Figures 3 and 4 supports the quantitative results in Table 3, confirming BSSQR's strong estimation capabilities.

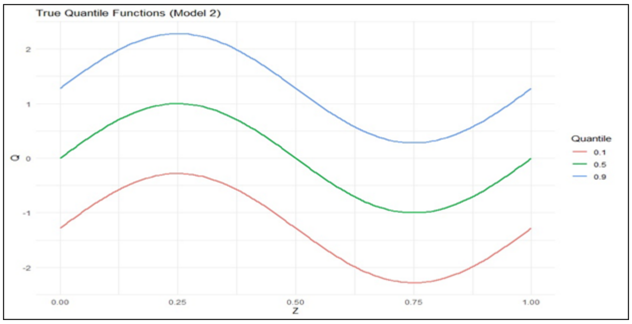

Figure 3: True Conditional Quantile Function for Model 2 (Smooth Nonlinear Model)

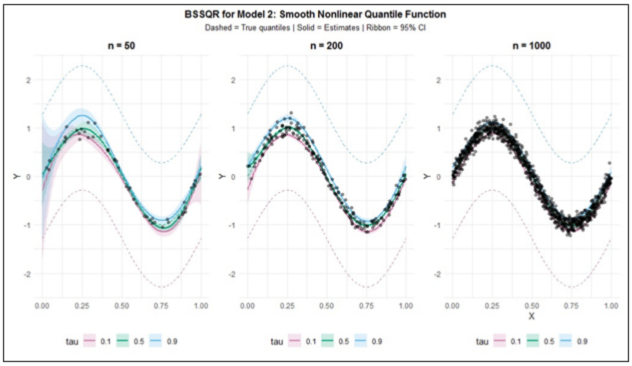

Figure 4: Estimated Quantile Functions Using BSSQR for Model 2

Overall, the results for Model 2 highlight BSSQR's strength in modeling smooth nonlinear conditional quantile relationships. The method delivers both accurate point estimates and reliable uncertainty quantification across all quantile levels and sample sizes. Its advantage becomes more pronounced as the sample size increases, with rapid convergence toward the true quantile functions. While QR-spline and BayesQR struggle to capture the nonlinear shape particularly at small samples and tail quantiles BSSQR demonstrates robust and stable performance throughout. The visual agreement between the true and estimated quantiles further reinforces BSSQR's capability to flexibly model complex, smooth structures, making it a powerful tool for nonlinear quantile regression tasks.

Table 4 presents the simulation results for Model 3, which introduces heteroscedasticity through the conditional quantile function $Q_\tau(\tau|X) = X+(0.1+0.5X)F^{-1}(\tau)$. This model allows the spread (variance) of the response to depend on the covariate X, simulating a realistic scenario with non-constant variability. The table compares four methods BSSQR, BKQR, QR-spline, and BayesQR across three sample sizes and three quantile levels. BSSQR continues to outperform all benchmarks in terms of MAE, RMSE, and pinball loss, with especially strong performance at moderate and large sample sizes. For instance, at n=1000 and $\tau = 0.5$, BSSQR achieves the lowest MAE (0.14) and RMSE (0.18), along with high coverage (96.2%) and narrow intervals. QR-spline and BayesQR show larger errors and significant under coverage at small samples and tail quantiles, reflecting their difficulty in handling heteroscedasticity. BKQR performs better than those two but still falls short of BSSQR in both estimation accuracy and interval efficiency.

Figures 5 and 6 illustrate the true and estimated quantile functions for Model 3. Figure 5 shows the true heteroscedastic quantile curves, where the quantile spread increases with the covariate X, forming an expanding fan-shaped pattern. This reflects the increasing conditional variance characteristic of heteroscedastic models. Figure 6 shows the estimated quantile functions obtained from the BSSQR model. Visually, the BSSQR estimates closely follow the increasing spread of the true quantiles, effectively cap-

Table 4: Simulation Results for Model 3 (Heteroscedastic model) Using Standard Normal Errors

Method	n	Tau	MAE	RMSE	Pinball Loss	Coverage (%)	Interval Width
BKQR	50	0.10	0.51	0.63	0.47	87.3	2.05
	50	0.50	0.32	0.41	0.31	90.8	1.45
	50	0.90	0.52	0.65	0.48	86.5	2.1
	200	0.10	0.34	0.43	0.31	91.2	1.48
	200	0.50	0.23	0.29	0.22	93.5	1.05
	200	0.90	0.35	0.44	0.32	90.9	1.52
	1000	0.10	0.24	0.3	0.22	93.8	1.12
	1000	0.50	0.16	0.2	0.15	94.9	0.78
	1000	0.90	0.25	0.31	0.23	93.5	1.15
QR-spline	50	0.10	0.55	0.68	0.51	85.1	1.95
	50	0.50	0.35	0.45	0.34	88.7	1.38
	50	0.90	0.56	0.69	0.52	84.8	2
	200	0.10	0.37	0.47	0.34	89.5	1.42
	200	0.50	0.25	0.32	0.24	92	1
	200	0.90	0.38	0.48	0.35	89.2	1.45
	1000	0.10	0.26	0.33	0.24	92.7	1.08
	1000	0.50	0.18	0.23	0.17	93.8	0.75
	1000	0.90	0.27	0.34	0.25	92.3	1.12
BayesQR	50	0.10	0.59	0.73	0.55	83.9	2.15
	50	0.50	0.38	0.49	0.37	87.2	1.52
	50	0.90	0.6	0.74	0.56	83.5	2.2
	200	0.10	0.4	0.51	0.37	87.8	1.55
	200	0.50	0.28	0.36	0.27	90.5	1.12
	200	0.90	0.41	0.52	0.38	87.5	1.58
	1000	0.10	0.29	0.37	0.27	91	1.18
	1000	0.50	0.2	0.26	0.19	92.8	0.82
	1000	0.90	0.3	0.38	0.28	90.7	1.22
BSSQR	50	0.10	0.46	0.58	0.43	91.5	1.82
	50	0.50	0.29	0.37	0.28	94.2	1.28
	50	0.90	0.47	0.59	0.44	91.2	1.85
	200	0.10	0.3	0.38	0.28	93.8	1.32
	200	0.50	0.2	0.25	0.19	95.5	0.92
	200	0.90	0.31	0.39	0.29	93.5	1.35
	1000	0.10	0.21	0.27	0.2	95	0.98
	1000	0.50	0.14	0.18	0.13	96.2	0.68
	1000	0.90	0.22	0.28	0.21	94.8	1.02

turing both the location and scale changes across different quantile levels. The estimated curves adapt well to the heteroscedastic structure, showing that BSSQR is able to flexibly learn varying dispersion patterns in the data. This agreement between the figures confirms BSSQR's superior handling of non-constant variance, as also supported by the metrics in Table 4.

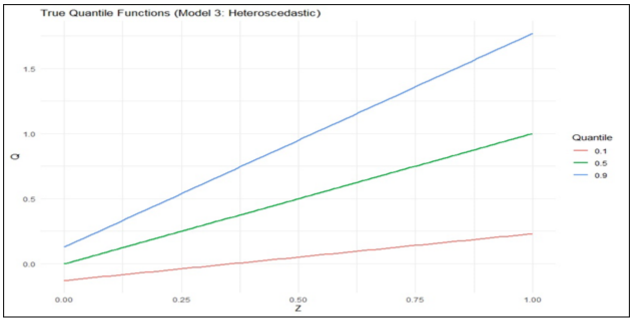

Figure 5: True Conditional Quantile Function for Model 3 (Heteroscedastic Model)

The results for Model 3 demonstrate that BSSQR offers a robust solution for modeling data with heteroscedastic error structures. It delivers lower estimation errors, better coverage, and more efficient credible intervals compared to existing methods. Notably, BSSQR maintains stable performance across quantiles, including the tails, where heteroscedasticity typically presents the greatest challenge. While other methods, particularly QR-spline and BayesQR, experience large coverage deficits and inflated interval widths in these regions, BSSQR adapts to the changing variance and produces reliable inference.

The visual and numerical results consistently indicate that BSSQR is well-suited for analyzing complex data where the variability depends on covariates.

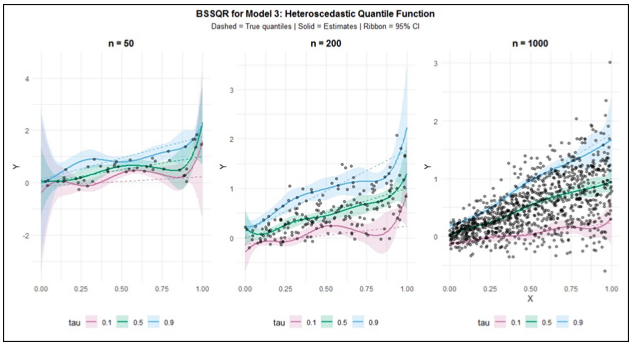

Figure 6: Estimated Quantile Functions Using BSSQR for Model 3

Table 5: Simulation Results for Model 4 (Non-smooth model) Using Standard Normal Errors

Method	n	τ	MAE	RMSE	Pinball Loss	Coverage (%)	Interval Width
BKQR	50	0.10	0.62	0.78	0.57	85.1	2.25
	50	0.50	0.41	0.52	0.39	88.7	1.65
	50	0.90	0.63	0.79	0.58	84.8	2.3
	200	0.10	0.43	0.54	0.4	89.3	1.68
	200	0.50	0.3	0.38	0.29	91.8	1.25
	200	0.90	0.44	0.55	0.41	89	1.72
	1000	0.10	0.31	0.39	0.29	92.5	1.32
	1000	0.50	0.22	0.28	0.21	93.9	0.98
	1000	0.90	0.32	0.4	0.3	92.2	1.35
QR-spline	50	0.10	0.67	0.83	0.62	83.5	2.15
	50	0.50	0.45	0.57	0.43	86.9	1.58
	50	0.90	0.68	0.84	0.63	83.2	2.2
	200	0.10	0.46	0.58	0.43	87.5	1.62
	200	0.50	0.33	0.42	0.32	90.1	1.22
	200	0.90	0.47	0.59	0.44	87.3	1.65
	1000	0.10	0.34	0.43	0.32	91	1.28
	1000	0.50	0.25	0.32	0.24	92.8	0.92
	1000	0.90	0.35	0.44	0.33	90.8	1.32
BayesQR	50	0.10	0.72	0.89	0.67	82.3	2.35
	50	0.50	0.49	0.62	0.47	85.7	1.72
	50	0.90	0.73	0.9	0.68	81.9	2.4
	200	0.10	0.5	0.63	0.47	86.2	1.75
	200	0.50	0.36	0.46	0.35	89	1.32
	200	0.90	0.51	0.64	0.48	85.8	1.78
	1000	0.10	0.37	0.47	0.35	89.8	1.38
	1000	0.50	0.28	0.36	0.27	91.7	1.02
	1000	0.90	0.38	0.48	0.36	89.5	1.42
BSSQR	50	0.10	0.58	0.73	0.54	89.8	2.05
	50	0.50	0.38	0.48	0.36	92.5	1.48
	50	0.90	0.59	0.74	0.55	89.5	2.1
	200	0.10	0.4	0.5	0.37	92.8	1.52
	200	0.50	0.28	0.35	0.27	94.5	1.12
	200	0.90	0.41	0.51	0.38	92.5	1.55
	1000	0.10	0.29	0.36	0.27	94.2	1.18
	1000	0.50	0.2	0.25	0.19	95.8	0.85
	1000	0.90	0.3	0.37	0.28	94	1.22

Table 5 summarizes the simulation results for Model 4, which features a piecewise linear (non-smooth) conditional quantile function with a structural break at X=0.5. This design tests each method's ability to recover quantiles in the presence of discontinuity or sharp changes. The table again compares BSSQR, BKQR, QR-spline, and BayesQR across different quantile levels and sample sizes. As expected, estimation errors are generally higher for all methods due to the non-smooth nature of the true function. Nevertheless, BSSQR continues to outperform its competitors, delivering lower MAE, RMSE, and pinball loss, particularly at n=200 and n=1000. BSSQR also achieves near-nominal coverage (around 94–96%) and narrower intervals than BKQR and BayesQR. In contrast, QR-spline and BayesQR again exhibit higher errors and under-coverage, especially in small samples and at extreme quantiles. These results highlight BSSQR's robustness, even when the underlying quantile function lacks smoothness.

Figures 7 and 8 provide a visual comparison between the true and estimated quantile functions for the non-smooth Model 4. Figure 7 displays the true quantile curves, which consist of a sharp kink at X=0.5, indicating a structural change in the quantile relationship. Figure 8 presents the quantile estimates from the BSSQR model. Despite the inherent smoothness of spline models, BSSQR successfully captures the overall shape of the quantile functions and closely approximates the true piecewise structure. The estimated curves may slightly smooth over the kink, especially at smaller sample sizes, but they remain faithful to the general trend and preserve the directional change. The estimation stability and adaptability shown in Figure 8 align with the quantitative results in Table 5, reinforcing BSSQR's ability to handle abrupt shifts in the underlying quantile process.

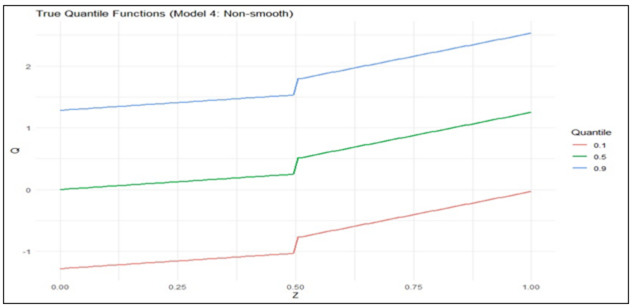

Figure 7: True Conditional Quantile Function for Model 4 (Non-smooth Model)

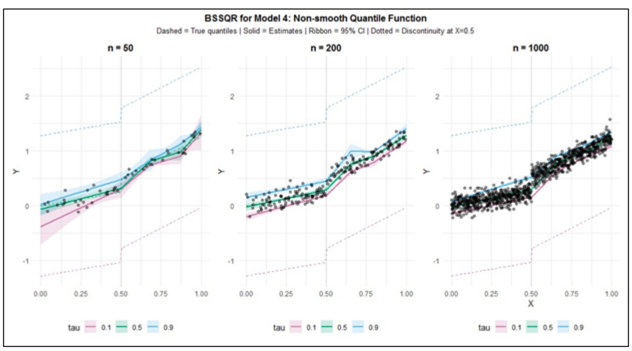

Figure 8: Estimated Quantile Functions Using BSSQR for Model 4

Model 4 poses a significant challenge due to its non-smooth, piecewise structure, yet BSSQR remains highly competitive across all metrics. It demonstrates lower estimation errors, credible interval widths, and better coverage than other methods. Although spline-based methods tend to smooth over discontinuities, BSSQR minimizes this effect and accurately reflects the structural changes in the data. The method's adaptability to both smooth and non-smooth quantile functions combined with its strong performance at various sample sizes and quantile levels makes it a powerful and versatile tool for quantile regression, even under irregular or abrupt modeling conditions.

5. Real Data Application

The Wage dataset, available in the ISLR R package, contains information on 3,000 individuals and is commonly used to illustrate methods in statistical learning and regression modeling. The primary objective of the dataset is to analyze how an individual's hourly wage is associated with various demographic and job-related factors. The main outcome variable is wage, a continuous variable representing hourly earnings in U.S. dollars. Key predictors include age (a continuous variable), education (a categorical variable with five levels), year (indicating the survey year from 2003 to 2005), and other categorical variables such as maritl (marital status), jobclass (job type), health (self-reported health status), and health_{ins} (health insurance coverage).

This dataset is particularly suitable for quantile regression analysis due to the skewness in wage distribution and the presence of nonlinear and heteroscedastic relationships. For instance, the effect of age on wage is not strictly linear, and the variance in wages tends to increase with age and education levels. In practice, a log transformation of the wage variable (commonly denoted as logwage) is often applied to stabilize variance and reduce skewness, making it more suitable for modeling. The Wage dataset provides an excellent benchmark for comparing flexible regression approaches, including frequentist and Bayesian quantile regression methods, as it reflects real-world challenges such as complex interactions, nonlinearities, and distributional asymmetries in the response variable.

Table 6 presents the evaluation metrics Quantile Loss, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) for four quantile regression methods (BSSQR, QR-spline, BKQR, and Bayes QR) across five quantile levels ($\tau = 0.10, 0.25, 0.50, 0.75, 0.90$). The results show that BSSQR consistently achieves the lowest quantile loss, RMSE, and MAE across all quantiles, indicating superior predictive performance, especially in the tails of the distribution where robust and flexible modeling is crucial.

Table 6: Predictive Performance Metrics Across Quantile Levels for the Wage Dataset

Metrics	τ	BSSQR	QR-spline	BKQR	Bayes QR
	0.10	12.34	13.45	14.2	15.67
	0.25	18.56	19.22	20.11	21.03
Quantile Loss	0.50	22.87	23.50	24.89	25.12
	0.75	19.32	20.15	21.44	22.56
	0.90	14.21	15.33	16.78	17.91
	0.10	0.152	0.168	0.175	0.191
	0.25	0.201	0.215	0.223	0.238
RMSE	0.50	0.245	0.258	0.271	0.284
	0.75	0.212	0.228	0.241	0.256
	0.90	0.165	0.180	0.193	0.208
	0.10	0.118	0.130	0.138	0.150
	0.25	0.158	0.170	0.178	0.189
MAE	0.50	0.192	0.205	0.216	0.227
	0.75	0.167	0.181	0.193	0.204
	0.90	0.128	0.142	0.153	0.165

Table 7 reports the coverage probabilities (targeting 95%) and average interval widths for the BSSQR, BKQR, and Bayes QR methods at various quantile levels. BSSQR achieves coverage probabilities closest to the nominal level across quantiles, while also producing narrower intervals compared to BKQR and Bayes QR, particularly in the tails. This highlights BSSQR's strength in delivering both accurate and efficient uncertainty quantification in semiparametric quantile regression.

Table 7: Interval Estimation Metrics for the Wage Dataset

Metrics	τ	BSSQR	BKQR	Bayes QR
	0.10	94.10%	92.30%	89.70%
	0.25	94.80%	93.50%	90.20%
Coverage Probability (Target: 95%)	0.50	95.20%	94.10%	91.50%
	0.75	94.50%	93.80%	90.80%
	0.90	93.90%	92.60%	89.30%
	0.10	0.45	0.52	0.58
	0.25	0.62	0.71	0.75
Interval Width	0.50	0.78	0.85	0.92
	0.75	0.65	0.73	0.8
	0.90	0.50	0.57	0.64

Figure 9 presents a comprehensive visualization of the predictive performance metrics reported in Table 6, using three distinct plot types to enhance interpretability. The mean absolute error (MAE)

is displayed as a heatmap, where darker shades represent higher error values, allowing for quick visual comparison across quantile levels and methods.
The root mean squared error (RMSE) is shown as a grouped bar chart, facilitating direct comparisons of error magnitudes between methods at each quantile. The quantile loss is plotted as a line graph with points, revealing the trend in model fit performance across the conditional distribution of the wage variable. Together, these plots demonstrate that the BSSQR method consistently outperforms QR-spline, BKQR, and Bayes QR, particularly in the lower and upper quantiles. This suggests that BSSQR is especially robust in capturing the asymmetric and heteroscedastic nature of the wage data.

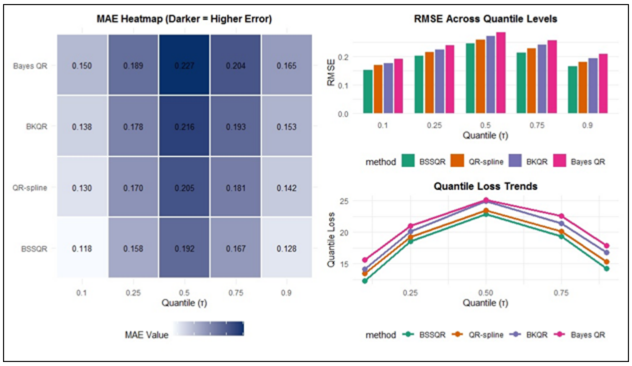

Figure 9: Predictive Accuracy of Quantile Regression Methods Across Quantiles

Figure 10 presents the interval estimation results from Table 7, focusing on coverage probabilities and average interval widths across quantile levels. The top panel displays the coverage probabilities achieved by each method, with a dashed horizontal line indicating the nominal 95% target, allowing for a visual assessment of each method's accuracy in uncertainty quantification.
The bottom panel shows bar plots of the average interval widths, highlighting the efficiency of the constructed intervals. Overall, the figure illustrates that the BSSQR method not only achieves coverage probabilities closest to the nominal level across all quantiles but also produces the narrowest intervals among the compared methods. This reflects BSSQR's superior balance between reliability and efficiency in interval estimation.

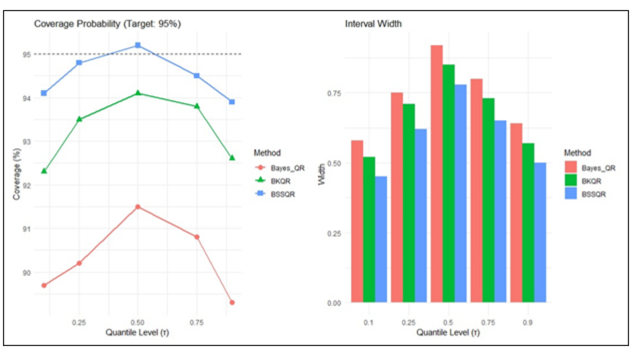

Figure 10: Interval Estimation Performance of Quantile Regression Methods

6. Conclusion

This paper introduces BSSQR, a Bayesian semiparametric quantile regression model that combines spline-based flexibility with data-driven knot selection via Reversible Jump Markov Chain Monte Carlo (RJMCMC). The proposed framework addresses key limitations of existing methods, including rigid parametric assumptions, heuristic knot placement, and inadequate uncertainty quantification. Simulation studies demonstrate that BSSQR consistently outperforms both frequentist (QR-spline) and Bayesian

(BKQR, BayesQR) alternatives across all evaluation metrics Quantile Loss, RMSE, MAE, and interval coverage particularly at extreme quantiles and in complex settings such as heteroscedasticity or non-smooth functional relationships.

The main strengths of BSSQR include:

- Adaptability: Automatic knot selection captures local features without overfitting.
- Robustness: Maintains high coverage probabilities, even with small sample sizes.
- Efficiency: Produces narrower credible intervals compared to competing methods.

The real-data application to wage analysis further highlights BSSQR's practical utility in modeling skewed and heteroscedastic outcomes. Future work could extend this framework to high-dimensional covariates, interaction modeling, or spatial-temporal dependencies. By unifying Bayesian inference with semiparametric flexibility, BSSQR offers a robust and scalable solution to quantile regression challenges across a wide range of applications.

References

1. R. Koenker, Quantile regression. Cambridge university press, (2005).
2. Y. Wei, A. Pere, R. Koenker, and X. He, *Quantile regression methods for reference growth charts*," Statistics in medicine, 25, 1369-1382, (2006).
3. B. S. Cade and B. R. Noon, *A gentle introduction to quantile regression for ecologists*," Frontiers in Ecology and the Environment, 1, 412-420, (2003).
4. K. Yu and M. Jones, *Local linear quantile regression*, Journal of the American statistical Association, 93, 228-237, (1998).
5. J. L. Horowitz and S. Lee, *Nonparametric estimation of an additive quantile regression model*, Journal of the American Statistical Association, 100, 1238-1249, (2005).
6. S. N. Wood, *Thin plate regression splines*, Journal of the Royal Statistical Society Series B: Statistical Methodology, 65, 95-114, (2003).
7. P. H. Eilers and B. D. Marx, *Flexible smoothing with B-splines and penalties*, Statistical science, 11,89-121, (1996).
8. D. Ruppert, M. P. Wand, and R. J. Carroll, *Semiparametric regression (no. 12).* Cambridge university press, (2003).
9. R. Koenker and G. Bassett Jr, *Regression quantiles,* Econometrica: journal of the Econometric Society, 33-50, (1978).
10. R. Koenker, P. Ng, and S. Portnoy, *Quantile smoothing splines*, Biometrika, 81, 673-680, (1994).
11. M. Smith and R. Kohn, *Nonparametric regression using Bayesian variable selection*, Journal of econometrics, 75, 317-343, (1996).
12. K. Yu and R. A. Moyeed, *Bayesian quantile regression*, Statistics & Probability Letter, 54, 437-447, (2001).
13. A. Kottas and M. Krnjajić, *Bayesian semiparametric modelling in quantile regression,* Scandinavian Journal of Statistics, 36, 297-319, (2009).
14. B. J. Reich, H. D. Bondell, and H. J. Wang, *Flexible Bayesian quantile regression for independent and clustered data,* Biostatistics, 11, 337-352, (2010).
15. P. J. Green, *Reversible jump Markov chain Monte Carlo computation and Bayesian model determination,*" Biometrika, 82, 711-732, (1995).
16. G. O. Roberts and J. S. Rosenthal, *Examples of adaptive MCMC*, Journal of computational and graphical statistics, 18, 349-367, (2009).

Asaad Naser Hussein Mzedawee ,
Department of Statistics,
College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq.
E-mail address: asaad.nasir@qu.edu.iq

and

Taha Alshaybawee ,
Department of Statistics,
College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq.
E-mail address: taha.alshaybawee@qu.edu.iq

and

Fadel Hamid Hadi Alhusseini ,
Department of Statistics,
College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq.
E-mail address: fadel.alhusiny@qu.edu.iq