

Article

High-Dimensional Variable Selection for Quantile Regression Based on Variational Bayesian Method

Dengluan Dai, Anmin Tang *  and Jinli Ye

Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, China; dd112021213126@mail.ynu.edu.cn (D.D.); jinliye@mail.ynu.edu.cn (J.Y.)

* Correspondence: tam1973as@ynu.edu.cn; Tel.: +86-187-87121838

Abstract: The quantile regression model is widely used in variable relationship research of moderate sized data, due to its strong robustness and more comprehensive description of response variable characteristics. With the increase of data size and data dimensions, there have been some studies on high-dimensional quantile regression under the classical statistical framework, including a high-efficiency frequency perspective; however, this comes at the cost of randomness quantification, or the use of a lower efficiency Bayesian method based on MCMC sampling. To overcome these problems, we propose high-dimensional quantile regression with a spike-and-slab lasso penalty based on variational Bayesian (VBSSLQR), which can, not only improve the computational efficiency, but also measure the randomness via variational distributions. Simulation studies and real data analysis illustrated that the proposed VBSSLQR method was superior or equivalent to other quantile and nonquantile regression methods (including Bayesian and non-Bayesian methods), and its efficiency was higher than any other method.

Keywords: quantile regression; spike-and-slab prior; variational Bayesian; high-dimensional data

MSC: 62-08



Citation: Dai, D.; Tang, A.; Ye, J. High-Dimensional Variable Selection for Quantile Regression Based on Variational Bayesian Method. *Mathematics* **2023**, *11*, 2232. <https://doi.org/10.3390/math11102232>

Academic Editors: Hongzhi Wang and Ye Chen

Received: 12 April 2023

Revised: 1 May 2023

Accepted: 8 May 2023

Published: 10 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Quantile regression, as introduced by Koenker and Bassett (1978) [1], is an important statistical inference about the relationship between quantiles of the response distribution and available covariates, and can offer a practically significant alternative to the traditional mean regression, because it provides a more comprehensive description of the response distribution than the mean. Moreover, quantile regression can capture the heterogeneous impact of regressors on different parts of the distribution [2], has excellent computational properties [3], exhibits robustness to outliers, and has a wide applicability [4]. For these reasons, quantile regression has attracted extensive attention in the literature. For example, see [5] for a Bayesian quantile regression with asymmetric Laplace distribution used to specify the likelihood, [6] for a Bayesian nonparametric approach to inference for quantile regression, ref. [7] for a mechanism of Bayesian inference of quantile regression models, and [8] for model selection in quantile regression, among others.

Although there is a growing literature on quantile regression, to the best of our knowledge, little of the existing quantile regression models have focused on high-dimensional data with variable numbers and larger sample sizes. In practice, a large number of variables may be collected and some of these will be insignificant and should be excluded from the final model. In the past two decades, there has been active methodological research on penalized methods for significant variable selection in linear parametric models. For example, see [9] for ridge regression, ref. [10] for a least absolute shrinkage and selection operator (Lasso), ref. [11] for smoothly clipped absolute deviation penalty (SCAD), ref. [12] for elastic net penalty, and ref. [13] for adaptive lasso methods. These methods have been extended to quantile regression; for example, see [14] for a L_1 -regularization method

for quantile regression, and ref. [15] for variable selection of quantile regression using SCAD and adaptive-Lasso penalties. Nevertheless, the aforementioned regularization methods discussed are both computationally complex and unstable. Additionally, they fail to account for prior information about parameters, which can lead to an unsatisfactory parametric estimation accuracy. In recent decades, Bayesian approaches to variable selection and parameter estimation have garnered significant attention. This is because they can substantially enhance the accuracy and efficiency of parametric estimation, by imposing various priors on model parameters, consistently selecting crucial variables, and providing more information for variable selection than penalization methods for highly nonconvex optimization problems. For example, see [16] for a Bayesian Lasso where the L_1 penalty is involved in Laplace prior, [17] for a Bayesian form of adaptive Lasso, ref. [18] for Bayesian Lasso quantile regression (BLQR), and [19] for Bayesian adaptive Lasso quantile regression (BALQR). The literature mentioned above implemented the standard Gibbs sampler for posterior computation, which is not easily scalable for high-dimensional data, where the number of variables is large compared with the sample size [20].

To address this issue, the Bayesian variable selection method with a spike-and-slab prior [20] has been favored by researchers, and this can be applied to high-dimensional data at the cost of a heavy computation burden. As a computationally efficient alternative to Markov chain Monte Carlo (MCMC) simulation, variational Bayes (VB) methods are gaining traction in machine learning and statistics for approximating posterior distributions in Bayesian inference. High-efficiency variational Bayesian spike-and-slab lasso (VBSSL) has been explored for certain high-dimensional models. Ray and Szabo (2022) [21] used a VBSSL method in a high-dimensional linear model, with regression coefficient's prior specified as a mixture of Laplace distribution and Dirac mass. Yi and Tang (2022) [22] used VBSSL technology in high-dimensional linear mixed models, with a interesting prior parameter as a mixture of two Laplace distributions. However, to the best of our knowledge, there has been little work done on a VBSSL method for quantile regression. Xi et al. (2016) [23] considered Bayesian variable selection for nonparametric quantile regression with a small variable dimension, in which a spike-and-slab prior was chosen as a mixture of the point mass at zero and normal distribution. In this paper, to reduce the computational burden and quantify the parametric uncertainty, we propose quantile regression with spike-and-slab lasso penalty based on variational Bayesian (VBSSLQR), in which the prior is a mixture of two Laplace distributions, with a smaller or larger variance, respectively.

The main contributions of this paper are as follows: First, our proposed VBSSLQR method can perform variable selection for high-dimensional quantile regression at a relatively low computational cost, and without the need for nonconvex optimization, while also avoiding the curse of dimensionality problem. Second, in contrast to mean regressions, our proposed quantile approach offers a more systematic strategy for analyzing how covariates impact the various quantiles of the response distribution. Third, in ultra-high-dimensional regression, the mean regression errors are frequently presumed to be subGaussian, which is not required in our setting.

The rest of the paper is organized as follows: In Section 2, for high-dimensional data, we propose an efficient quantile regression with a spike-and-slab lasso penalty based on variational Bayes (VBSSLQR). In Section 3, we randomly generate high-dimensional data with $n = 200$ and $p = 500$ (excluding intercept items), and perform 500 simulation experiments, to explore the performance of our algorithm and compare it with other quantile regression methods (Bayesian and non-Bayesian) and nonquantile regression methods. The results show that our method is superior to other approaches in the case of high-dimensional data. We applied VBSSLQR to a real dataset that contained information about crime in various cities in the United States, and compared it with other quantile regression methods. The results showed that our method also had a good performance and excellent efficiency with real data, and the relevant results are shown in Section 4. Some concluding remarks are given in Section 5. Technical details are presented in the Appendices A–D.

2. Models and Methods

2.1. Quantile Regression

Consider a dataset of n independent subjects. For the i th subject, let y_i be the response, while $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ir})^\top$ is an $(r + 1) \times 1$ predictor vector, a simple linear regression model is defined as follows:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, i = 1, 2, \dots, n, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_r)^\top$ is the regression coefficient vector with β_0 corresponding to the intercept terms, and ε_i represents the error term with unknown distribution. It is usual to assume that the τ th quantile of the random error term is 0; that is, $Q_\tau(\varepsilon_i) = 0$ for $0 < \tau < 1$. According to this assumption, the form of τ th quantile regression of model (1) is specified as follows:

$$Q_{y_i}(\tau | \mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \tag{2}$$

where $Q_{y_i}(\tau | \mathbf{x}_i)$ is the inverse cumulative distribution function of y_i , given \mathbf{x}_i evaluated at τ . The estimate of the regression coefficient vector $\boldsymbol{\beta}$ in Equation (2) is

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \operatorname{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^{r+1}} \left[\sum_{i: y_i \geq \mathbf{x}_i^\top \boldsymbol{\beta}} \tau |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}| + \sum_{i: y_i < \mathbf{x}_i^\top \boldsymbol{\beta}} (1 - \tau) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}| \right] \\ &= \operatorname{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^{r+1}} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \end{aligned} \tag{3}$$

where the loss function $\rho_\tau(\mu) = \mu \times (\tau - I(\mu < 0))$ with the indicator function $I(\cdot)$.

In light of [5,24], minimizing Equation (3) is equivalent to maximizing the likelihood of n independent individuals with the i th one distributed as an asymmetric Laplace distribution (ALD) specified as

$$p(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \sigma, \tau) = \frac{\tau(1 - \tau)}{\sigma} \exp \left\{ -\rho_\tau \left(\frac{y_i - \mu_i}{\sigma} \right) \right\}, \tag{4}$$

where the local parameter $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, the scale parameter $\sigma > 0$, and the skewness parameter τ is between 0 and 1; obviously, ALD is a Laplace distribution when $\tau = 0.5$ and $\mu_i = 0$. However, it is computationally infeasible to carry out statistical inference based directly on Equation (4) involving the nondifferentiable point μ_i . Following [25], Equation (4) can be rewritten in the following hierarchical fashion:

$$\begin{cases} y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + k_1 z_i + \sqrt{k_2 \sigma z_i} \zeta_i, \\ z_i | \sigma \stackrel{\text{i.i.d.}}{\sim} \text{Exp} \left(\frac{1}{\sigma} \right), \\ \zeta_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1), \\ z_i \text{ is independent of } \zeta_i, \end{cases} \tag{5}$$

where $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $k_2 = \frac{2}{\tau(1-\tau)}$ and $\text{Exp} \left(\frac{1}{\sigma} \right)$ denotes the exponential distribution with mean σ , whose specific density function is $p(z_i | \sigma) = \frac{1}{\sigma} \exp \left(-\frac{1}{\sigma} z_i \right) I(z_i > 0)$. Equation (5) illustrates that an asymmetric Laplace distribution can also be represented as a mixture of exponential and standard normal distributions, which allows us to express a quantile regression model as a normal regression model, in which the response has the following conditional distribution:

$$y_i | \mathbf{x}_i, \boldsymbol{\beta}, z_i, \sigma \stackrel{\text{ind}}{\sim} N \left(\mathbf{x}_i^\top \boldsymbol{\beta} + k_1 z_i, k_2 \sigma z_i \right).$$

For the above-defined quantile regression model with high-dimensional covariate vector (r is large enough), it is of interest to estimate the parameter vector $\boldsymbol{\beta}$ and to identify

the critical covariates. To this end, we considered Bayesian quantile regression based on spike-and-slab lasso, as follows:

2.2. Bayesian Quantile Regression Based on a Spike-and-Slab Lasso

As early as 2016, Xi et al. [23] applied a spike-and-slab prior to Bayesian quantile regression, but their proposed prior was a mixture of zero particle and normal distribution with large variance, and the estimate of posterior density was obtained using a Gibbs sampler. To offer novel theoretical insights into a class of continuous spike-and-slab priors, Rockova (2018) [26] introduced a novel family of spike-and-slab priors, which are a mixture of two density functions with spike or slab probability. In this paper, we adopt a spike-and-slab lasso prior with a mixture of two Laplace distributions with large or small variance, respectively [26], which facilitates the variational Bayesian technique for approximating the posterior density of parameters and for improving the efficiency of the algorithm. In light of ref. [26], given the indicator $\gamma_j = 0$ or 1 , the prior of β in the Bayesian quantile regression model (5) can be written as

$$\pi(\beta|\gamma) = \prod_{j=0}^r \pi(\beta_j|\gamma_j) = \prod_{j=0}^r [\gamma_j \Psi_0(\beta_j|\lambda_0) + (1 - \gamma_j) \Psi_1(\beta_j|\lambda_1)], \tag{6}$$

where the Laplace density $\Psi_0(\beta_j|\lambda_0) = \frac{\lambda_0}{2} \exp(-\lambda_0|\beta_j|)$ and $\Psi_1(\beta_j|\lambda_1) = \frac{\lambda_1}{2} \exp(-\lambda_1|\beta_j|)$ with precision parameters λ_0 and λ_1 satisfying that $\lambda_0 \gg \lambda_1$, the indicator variable set $\gamma = \{\gamma_j|j = 0, 1, \dots, r\}$, the j th variable is active when $\gamma_j = 0$, and inactive otherwise. Similarly to [27], the Laplace distribution for the regression coefficient β_j can be represented as a mixture of a normal distribution and an exponential distribution, specifically the distribution of β_j can be expressed as a hierarchical structure, as follows:

$$\begin{aligned} \beta_j | h_{0j}^2, h_{1j}^2, \gamma_j &\stackrel{\text{i.i.d.}}{\sim} \gamma_j N(0, h_{0j}^2) + (1 - \gamma_j) N(0, h_{1j}^2), \\ h_{0j}^2 | \lambda_0^2 &\stackrel{\text{i.i.d.}}{\sim} \text{Exp}\left(\frac{\lambda_0^2}{2}\right), \\ h_{1j}^2 | \lambda_1^2 &\stackrel{\text{i.i.d.}}{\sim} \text{Exp}\left(\frac{\lambda_1^2}{2}\right), \\ \gamma_j &\stackrel{\text{i.i.d.}}{\sim} B(\pi_\gamma), \end{aligned} \tag{7}$$

where $B(\pi_\gamma)$ denotes the Bernoulli distribution, with π_γ being the probability that indicator variable γ_j equals one for $j = 0, 1, \dots, r$, and specifies the prior of π_γ as a Beta distribution $\text{Be}(a_{\pi_\gamma}, b_{\pi_\gamma})$, where hyperparameters a_{π_γ} and b_{π_γ} , λ_0^2 and λ_1^2 are regularization parameters to identify important variables, for which we consider the following conjugate priors:

$$\lambda_0^2 \sim \text{Ga}(v_{\lambda_0}, 1), \lambda_1^2 \sim \text{Ga}(v_{\lambda_1}, 1),$$

where $\text{Ga}(a, b)$ denotes the gamma distribution with shape parameter a and scale parameter b . As mentioned above, λ_0 and λ_1 should satisfy that $\lambda_0 \gg \lambda_1$, to this end, we select hyperparameters v_{λ_0} and v_{λ_1} to satisfy that $v_{\lambda_0} \gg v_{\lambda_1}$. The prior of scale parameter σ in (5) is an inverse gamma distribution $\text{IG}(a_\sigma, b_\sigma)$ of the hyperparameters $a_\sigma = 1$ and $b_\sigma = 0.01$ in the paper leading to almost non-informative prior.

Under a Bayesian statistical paradise, based on the above priors and likelihood of quantile regression, it is required to induce the following posterior distribution $\pi(\theta|\mathbf{D}) \propto p(\theta, \mathbf{D})$, where $\theta = \{\beta, z, \sigma, \gamma, h_0^2, h_1^2, \lambda_0^2, \lambda_1^2, \pi_\gamma\}$, the latent variable set $z = \{z_i|i = 1, \dots, n\}$, $h_0^2 = \{h_{0j}^2|j = 0, 1, \dots, r\}$, $h_1^2 = \{h_{1j}^2|j = 0, 1, \dots, r\}$, observing set $\mathbf{D} = \{\mathbf{y}, \mathbf{x}\}$ with the response set $\mathbf{y} = \{y_i|i = 1, \dots, n\}$ and covariate set $\mathbf{x} = \{x_i|i = 1, \dots, n\}$. Based on the hierarchical structure (5) of the quantile regression likelihood $p(y_i|x_i, \beta)$ and the hierarchical

structure (7) of the spike-and-slab prior to regression coefficient vector β , we derive the joint density

$$\begin{aligned}
 p(\theta, \mathbf{D}) &= \prod_{i=1}^n N(y_i | \mathbf{x}_i^\top \beta + k_1 z_i, k_2 \sigma z_i) \text{Exp}(z_i | \sigma^{-1}) \text{Ga}(\lambda_0^2 | \nu_{\lambda_0}, 1) \text{Ga}(\lambda_1^2 | \nu_{\lambda_1}, 1) \\
 &\times \prod_{j=0}^r [N(\beta_j | 0, h_{0j}^2)]^{\gamma_j} [N(\beta_j | 0, h_{1j}^2)]^{(1-\gamma_j)} \text{Exp}(h_{0j}^2 | \frac{\lambda_0^2}{2}) \text{Exp}(h_{1j}^2 | \frac{\lambda_1^2}{2}) \\
 &\times \prod_{j=0}^r B(\gamma_j | 1, \pi_\gamma) \text{Be}(\pi_\gamma | a_{\pi_\gamma}, b_{\pi_\gamma}) \text{IG}(\sigma | a_\sigma, b_\sigma) \\
 &\propto \prod_{i=1}^n (k_2 \sigma z_i)^{-\frac{1}{2}} \exp\left(-\frac{(y_i - \mathbf{x}_i^\top \beta - k_1 z_i)^2}{2k_2 \sigma z_i}\right) \frac{1}{\sigma} \exp\left(-\frac{z_i}{\sigma}\right) (\lambda_0^2)^{\nu_{\lambda_0}-1} \exp(-\lambda_0^2) \\
 &\times (\lambda_1^2)^{(\nu_{\lambda_1}-1)} \exp(-\lambda_1^2) \prod_{j=0}^r \left[\frac{1}{\sqrt{h_{0j}^2}} \exp\left(-\frac{\beta_j^2}{2h_{0j}^2}\right)\right]^{\gamma_j} \left[\frac{1}{\sqrt{h_{1j}^2}} \exp\left(-\frac{\beta_j^2}{2h_{1j}^2}\right)\right]^{(1-\gamma_j)} \\
 &\times \frac{\lambda_0^2}{2} \exp\left(-\frac{\lambda_0^2}{2} h_{0j}^2\right) \frac{\lambda_1^2}{2} \exp\left(-\frac{\lambda_1^2}{2} h_{1j}^2\right) \prod_{j=0}^r \pi_\gamma^{\gamma_j} (1 - \pi_\gamma)^{1-\gamma_j} \pi_\gamma^{a_{\pi_\gamma}-1} (1 - \pi_\gamma)^{b_{\pi_\gamma}-1} \\
 &\times \frac{b_\sigma^{a_\sigma}}{\Gamma(a_\sigma)} \sigma^{-a_\sigma-1} \exp\left(-\frac{b_\sigma}{\sigma}\right).
 \end{aligned} \tag{8}$$

Although sampling from the aforementioned posterior is simple, it becomes increasingly time-consuming for higher dimension quantile models. To tackle this issue, we developed a faster and more efficient alternative method based on variational Bayesian.

2.3. Quantile Regression with a Spike-and-Slab Lasso Penalty Based on Variational BAYESIAN

At present, the most commonly variational Bayesian approximation posterior distribution methods use mean field approximation theory [28], which has the highest efficiency among variational methods, especially for those parameters or parameter block with conjugate priors. Bayesian quantile regression needs to take into account that the variance of each observation value is different, and each y_i corresponds to a potential variable z_i , which will result in the algorithm efficiency of quantile regression being lower than that of general mean regression. Therefore, in this paper, we use the variational Bayesian algorithm of the mean field approximation, which is the most efficient algorithm, to derive the quantile regression model with the spike-and-slab lasso penalty.

Based on variational theory, we choose a densities for random variables θ from variational family \mathcal{F} , which having the same support Θ as the posterior density $\pi(\theta | \mathbf{D})$. We approximate the posterior density $\pi(\theta | \mathbf{D})$ by any variational density $q(\theta) \in \mathcal{F}$. The variational Bayesian method is to seek the optimal approximation to $\pi(\theta | \mathbf{D})$ by minimizing the Kullback-Leibler divergence between $q(\theta)$ and $\pi(\theta | \mathbf{D})$, which is an optimization problem that can be expressed as:

$$q^*(\theta) = \underset{q(\theta) \in \mathcal{F}}{\text{argmin}} \text{KL}[q(\theta) \| \pi(\theta | \mathbf{D})],$$

where $\text{KL}[q(\theta) \| \pi(\theta | \mathbf{D})] = \int_{\Theta} q(\theta) \log \frac{q(\theta)}{\pi(\theta | \mathbf{D})} d\theta$, which is not less than zero and equal to zero if, and only if, $q(\theta) \equiv \pi(\theta | \mathbf{D})$. The posterior density $\pi(\theta | \mathbf{D}) = \frac{p(\theta, \mathbf{D})}{p(\mathbf{D})}$ with the joint distribution $p(\theta, \mathbf{D})$ of parameter θ and data \mathbf{D} and the marginal distribution $p(\mathbf{D})$ data \mathbf{D} . Since $p(\mathbf{D}) = \int_{\Theta} p(\theta, \mathbf{D}) d\theta$, which does not have an analytic expression for our considered model, it is rather difficult to implement the optimization problem presented above. It is easy to induce that

$$\log p(\mathbf{D}) = \text{KL}[q(\theta) \| \pi(\theta | \mathbf{D})] + \mathcal{L}\{q(\theta)\} \geq \mathcal{L}\{q(\theta)\},$$

in which the evidence lower bound (ELOB) $\mathcal{L}\{q(\theta)\} = E_{q(\theta)}[\log p(\theta, D)] - E_{q(\theta)}[\log q(\theta)]$ with $E_{q(\theta)}[\cdot]$ representing the expectation taken with respect to the variational density $q(\theta)$. Thus, minimizing $KL[q(\theta) \parallel \pi(\theta | D)]$ is equivalent to maximizing $\mathcal{L}\{q(\theta)\}$ because $\log p(D)$ does not depend on θ . That is,

$$q^*(\theta) = \underset{q(\theta) \in \mathcal{F}}{\operatorname{argmin}} KL[q(\theta) \parallel \pi(\theta | D)] = \underset{q(\theta) \in \mathcal{F}}{\operatorname{argmax}} \mathcal{L}\{q(\theta)\},$$

which indicates that seeking the optimal approximation problem of $\pi(\theta | D)$ becomes maximizing $\mathcal{L}\{q(\theta)\}$ under the variational family \mathcal{F} . The complexity of the approximation problem heavily is related to the variational family \mathcal{F} . Therefore, choosing a comparatively simple variational family \mathcal{F} to optimize the objective function $\mathcal{L}\{q(\theta)\}$ with respect to $q(\theta)$ is fascinating.

Following the commonly used approach to choosing a tractable variational family \mathcal{F} in the variational studies, we consider the frequently-used mean-field theory, which assumes that blocks of θ are mutually independent and each is measured by the parameters of the variational density. Obviously, the variational density $q(\theta)$ is assumed to be factorized across the blocks of θ :

$$q(\theta) = \prod_{j=0}^r q_1(\beta_j, \gamma_j) q_2(h_{0j}^2) q_3(h_{1j}^2) \prod_{i=1}^n q_4(z_i) q_5(\lambda_0^2) q_6(\lambda_1^2) q_7(\pi_\gamma) q_8(\sigma) \equiv \prod_s^8 q_s(\theta_s),$$

in which form of each variational densities $q_s(\theta_s)$'s is unknown, but the above assumed factorization across components is predetermined. Moreover, the best solutions for $q_s(\theta_s)$'s are to be achieved by maximizing $\mathcal{L}\{q(\theta_1, \dots, \theta_8)\}$ with respect to variational densities $q_1(\theta_1), \dots, q_8(\theta_8)$ by the coordinate ascent method, where $\theta = \{\theta_1, \dots, \theta_8\}$ where θ_s can be either a scalar or a vector. This means that when the correlation between several unknown parameters or potential variables cannot be ignored, they should be put in the same block and merged into θ_s . Following the idea of the coordinate ascent method given in ref. [29], when fixing other variational factors $q_k(\theta_k)$ for $k \neq s$, i.e., the optimal density $q_s^*(\theta_s)$, which maximizes $\mathcal{L}\{q(\theta)\}$ with respect to $q_s(\theta_s)$, is shown to take the form

$$q_s^*(\theta_s) \propto \exp\{E_{-\theta_s}[\log p(\theta, D)]\}, \tag{9}$$

where $\log p(\theta, D)$ is the logarithm of the joint density function and $E_{-\theta_s}[\cdot]$ is the expectation taken with respect to the density $\prod_{k \neq s} q_k(\theta_k)$ for $s = 1, \dots, 8$.

According to Equations (8) and (9), we can derive the variational posterior for each parameter as follows (see Appendix A for the details):

$$\begin{aligned} \beta_j &\stackrel{\text{i.i.d.}}{\sim} \mu_{\gamma_j} N(\mu_{0j}, \sigma_{0j}^2) + (1 - \mu_{\gamma_j}) N(\mu_{1j}, \sigma_{1j}^2), \text{ for } j = 1, \dots, r, \\ h_{0j}^2 &\stackrel{\text{i.i.d.}}{\sim} \mu_{\gamma_j} \text{GIG}\left(\frac{1}{2}, E_{\lambda_0^2}(\lambda_0^2), \sigma_{0j}^2 + \mu_{0j}^2\right) + (1 - \mu_{\gamma_j}) \text{Exp}\left(\frac{1}{2} E_{\lambda_0^2}(\lambda_0^2)\right), \\ h_{1j}^2 &\stackrel{\text{i.i.d.}}{\sim} (1 - \mu_{\gamma_j}) \text{GIG}\left(\frac{1}{2}, E_{\lambda_1^2}(\lambda_1^2), \sigma_{1j}^2 + \mu_{1j}^2\right) + \mu_{\gamma_j} \text{Exp}\left(\frac{1}{2} E_{\lambda_1^2}(\lambda_1^2)\right), \\ \lambda_0^2 &\sim \text{Ga}\left(r + 1 + \nu_{\lambda_0}, 1 + \frac{1}{2} \sum_{j=1}^r E_{h_{0j}^2}(h_{0j}^2)\right), \\ \lambda_1^2 &\sim \text{Ga}\left(r + 1 + \nu_{\lambda_1}, 1 + \frac{1}{2} \sum_{j=1}^r E_{h_{1j}^2}(h_{1j}^2)\right), \\ \gamma_j &\stackrel{\text{i.i.d.}}{\sim} B(1, (1 + e^{\zeta_j})^{-1}), \text{ for } j = 0, 1, \dots, r, \\ \pi_\gamma &\sim \text{Be}\left(a + \sum_{j=0}^r \mu_{\gamma_j}, r + 1 + b - \sum_{j=0}^r \mu_{\gamma_j}\right), \end{aligned}$$

$$z_i \stackrel{\text{i.i.d.}}{\sim} \text{GIG}\left(\frac{1}{2}, a_{z_i}, b_{z_i}\right), \text{ for } i = 1, 2, \dots, n,$$

$$\sigma \sim \text{IG}\left(\frac{3}{2}n + a_\sigma, c_\sigma\right),$$

where $\text{GIG}(\cdot, \cdot, \cdot)$ denotes generalized inverse Gaussian distribution,

$$\begin{aligned} \mu_{0j} &= \sigma_{0j}^2 \frac{\mathbb{E}_\sigma(\sigma^{-1})}{k_2} \sum_{i=1}^n \left[\mathbb{E}_{z_i}(z_i^{-1}) \left(y_i - \mathbf{x}_{i,(-j)}^\top \boldsymbol{\mu}_{\beta_{(-j)}} \right) - k_1 \right] x_{ij}, \\ \sigma_{0j}^2 &= \left[\mathbb{E}_{h_{0j}^2 | \gamma_j=1} (h_{0j}^{-2}) + \frac{\mathbb{E}_\sigma(\sigma^{-1})}{k_2} \sum_{i=1}^n x_{ij}^2 \mathbb{E}_{z_i}(z_i^{-1}) \right]^{-1}, \\ \mu_{1j} &= \sigma_{1j}^2 \frac{\mathbb{E}_\sigma(\sigma^{-1})}{k_2} \sum_{i=1}^n \left[\mathbb{E}_{z_i}(z_i^{-1}) \left(y_i - \mathbf{x}_{i,(-j)}^\top \boldsymbol{\mu}_{\beta_{(-j)}} \right) - k_1 \right] x_{ij}, \\ \sigma_{1j}^2 &= \left[\mathbb{E}_{h_{1j}^2 | \gamma_j=0} (h_{1j}^{-2}) + \frac{\mathbb{E}_\sigma(\sigma^{-1})}{k_2} \sum_{i=1}^n x_{ij}^2 \mathbb{E}_{z_i}(z_i^{-1}) \right]^{-1}, \end{aligned} \tag{10}$$

and

$$\begin{aligned} \zeta_j &= \mathbb{E}_{\pi_\gamma}(\log(1 - \pi_\gamma)) - \mathbb{E}_{\pi_\gamma}(\log \pi_\gamma) + \frac{1}{2} \mathbb{E}_{h_{0j}^2 | \gamma_j=1}(\log h_{0j}^2) - \frac{1}{2} \mathbb{E}_{h_{1j}^2 | \gamma_j=0}(\log h_{1j}^2) \\ &\quad + \frac{1}{2} \{\log \sigma_{1j}^2 - \log \sigma_{0j}^2\} + \frac{1}{2} (\log \sigma_{1j}^2 - \log \sigma_{0j}^2) + \frac{1}{2} (\mu_{1j}^2 (\sigma_{1j}^2)^{-1} - \mu_{0j}^2 (\sigma_{0j}^2)^{-1}), \\ a_{z_i} &= \frac{\mathbb{E}_\sigma(\sigma^{-1})}{k_2} (k_1^2 + 2k_2), \\ b_{z_i} &= \frac{\mathbb{E}_\sigma(\sigma^{-1})}{k_2} [(y_i - \mathbf{x}_i^\top \boldsymbol{\mu}_\beta)^2 + \mathbf{x}_i^\top \boldsymbol{\Sigma}_\beta \mathbf{x}_i], \\ c_\sigma &= \frac{1}{2k_2} \sum_{i=1}^n \{k_1^2 \mu_{z_i} + \mathbb{E}_{z_i}(z_i^{-1}) [(y_i - \mathbf{x}_i^\top \boldsymbol{\mu}_\beta)^2 + \mathbf{x}_i^\top \boldsymbol{\Sigma}_\beta \mathbf{x}_i] \\ &\quad - 2k_1 (y_i - \mathbf{x}_i^\top \boldsymbol{\mu}_\beta)\} + \sum_{i=1}^n \mu_{z_i} + b_\sigma. \end{aligned}$$

In the above equation, $\mu_{\lambda_0^2} = \mathbb{E}_{\lambda_0^2}(\lambda_0^2)$, and $\mu_{\lambda_1^2}, \mu_{\beta_j}, \mu_{\gamma_j}, \mu_{z_i}$ are similar to $\mu_{\lambda_0^2}$, $\boldsymbol{\mu}_\beta = (\mu_{\beta_0}, \mu_{\beta_1}, \dots, \mu_{\beta_p})^\top$ with $\mu_{\beta_j} = \mu_{\gamma_j} \mu_{0j} + (1 - \mu_{\gamma_j}) \mu_{1j}$, and $\boldsymbol{\mu}_{\beta_{(-j)}}$ is a $p \times 1$ vector with the j th component of vector $\boldsymbol{\mu}_\beta$ deleted, $\boldsymbol{\Sigma}_\beta = \text{diag}(\sigma_{\beta_0}^2, \sigma_{\beta_1}^2, \dots, \sigma_{\beta_p}^2)$ with $\sigma_{\beta_j}^2 = \mu_{\gamma_j} \sigma_{0j}^2 + (1 - \mu_{\gamma_j}) \sigma_{1j}^2 + \mu_{\gamma_j} (1 - \mu_{\gamma_j}) (\mu_{0j} - \mu_{1j})^2$.

In the section above, we derived the variational posterior of each parameter. Using the idea of coordinate axis optimization, we can update each variational distribution iteratively until it converges.

For this reason, we list the variational Bayesian spike-and-slab lasso quantile regression (VBSSLQR) algorithm as shown in Algorithm 1:

Algorithm 1 Variational Bayesian spike-and-slab lasso quantile regression (VBSSLQR).

Input:

Data y , predictors x , prior parameters $\nu_{\lambda_0} = 10^4, \nu_{\lambda_1} = 1, a_\sigma = 1, b_\sigma = 0.01, a = b = 1$, precision $\epsilon = 0.01$ and quantile τ ;

Output:

Optimized variational parameters $E_{\beta_j}(\beta_j), \text{Var}_{\beta_j}(\beta_j)$ for $j = 0, 1, \dots, r$ and the corresponding Bayesian confidence interval.

Initialize: $\mu_\beta^{(0)}; E_{h_{0j}^2|\gamma_j=1}(\frac{1}{h_{0j}^2}) = 0.01; E_{h_{1j}^2|\gamma_j=0}(\frac{1}{h_{1j}^2}) = 1; \mu_\gamma^{(0)} = \mathbf{0.5};$

$E_{\lambda_0^2}^{(0)}(\lambda_0^2) = 100; E_{\lambda_1^2}^{(0)}(\lambda_1^2) = 1; E_{\pi_\gamma}^{(0)}(\log \pi_\gamma) = 0; E_{\sigma}^{(0)}(\sigma^{-1}) = 1;$

$E_{\pi_\gamma}^{(0)}(\log(1 - \pi_\gamma)) = -1; \gamma^{(0)} = \mathbf{0}; E_{z_i}^{(0)}(z_i) = E_{z_i}^{(0)}(z_i^{-1});$

while $|d^{(t)}| > \epsilon$ **do**

for $j = 0$ to r **do**

 Update $\sigma_{0j}^{2(t+1)}, \mu_{0j}^{(t+1)}, \sigma_{1j}^{2(t+1)}$ and $\mu_{1j}^{(t+1)}$ according to Equation (10).

 Update $E_{h_{0j}^2|\gamma_j=1}^{(t+1)}(h_{0j}^2), E_{h_{0j}^2}^{(t+1)}(h_{0j}^2), E_{h_{0j}^2|\gamma_j=1}^{(t+1)}(h_{0j}^{-2}), E_{h_{0j}^2|\gamma_j=1}^{(t+1)}(\log h_{0j}^2)$ according to $q(h_{0j}^2)$,

 Update $E_{h_{1j}^2|\gamma_j=0}^{(t+1)}(h_{1j}^2), E_{h_{1j}^2}^{(t+1)}(h_{1j}^2), E_{h_{1j}^2|\gamma_j=0}^{(t+1)}(h_{1j}^{-2}), E_{h_{1j}^2|\gamma_j=0}^{(t+1)}(\log h_{1j}^2)$ according to $q(h_{1j}^2)$,

 Update $E_{\gamma_j}^{(t+1)}(\gamma_j)$ according to the variational posterior $q(\gamma_j)$,

 Update $E_{\beta_j}^{(t+1)}(\beta_j)$ and $\text{Var}_{\beta_j}^{(t+1)}(\beta_j)$,

end for

 Update $E_{\lambda_0^2}^{(t+1)}(\lambda_0^2)$ according to the variational posterior $q(\lambda_0^2)$,

 Update $E_{\lambda_1^2}^{(t+1)}(\lambda_1^2)$ according to the variational posterior $q(\lambda_1^2)$,

 Update $E_{\pi_\gamma}^{(t+1)}(\pi_\gamma), E_{\pi_\gamma}^{(t+1)}(\log \pi_\gamma)$ and $E_{\pi_\gamma}^{(t+1)}(\log(1 - \pi_\gamma))$ according to $q(\pi_\gamma)$,

for $i = 1$ to n **do**

 Update $E_{z_i}^{(t+1)}(z_i)$ and $E_{z_i}^{(t+1)}(z_i^{-1})$ according to the variational posterior $q(z_i)$,

end for

 Update $E_{\sigma}^{(t+1)}(\sigma)$ and $E_{\sigma}^{(t+1)}(\sigma^{-1})$ according to the variational posterior $q(\sigma)$,

$|d^{(t+1)}| = \max\{|\theta_{q_1}^{(t+1)} - \theta_{q_1}^{(t)}|, \dots, |\theta_{q_m}^{(t+1)} - \theta_{q_m}^{(t)}|\}$,

end while

In Algorithm 1 above, $\Psi(\cdot)$ is the digamma function and the expectation $E_{h_{0j}^2|\gamma_j=1}^{(t+1)}(\log h_{0j}^2)$ of $\log h_{0j}^2$ with respect to generalized inverse Gaussian distribution. Thus, we assume that $x \sim \text{GIG}(p, a, b)$, then:

$$E_x(\log x) = \frac{dK_p(\sqrt{ab})}{K_p(\sqrt{ab})} - \frac{1}{2} \ln\left(\frac{a}{b}\right), \tag{11}$$

where $K_p(\cdot)$ represents the Bessel function of the second kind. Note that there is no analytic solution or function to the differential of the modified Bessel function. Therefore, we approximate $E_x(\log x)$ using the second-order Taylor expansion of $\log x$. This paper lists the expectations of some parameter functions about variational posteriors involved in Algorithm 1; see Appendix B for details. Based on our proposed VBSSLQR algorithm, in the next section, we randomly generate high-dimensional data, conduct simulation research, and compare the performance with other methods. Notably, the asymptotic variance of the quantile regression is reciprocal proportional to the density of the errors at the quantile point. In cases where n is small and we estimate extreme quantiles, the correlating asymptotic variance will be large, resulting in less precise estimates [23]. Therefore, the regression coefficient is difficult to estimate at an extreme quantile and this is feasible when the sample sizes needs to be increased appropriately.

3. Simulation Studies

In this section, we used simulated high-dimensional data with a sample size $n = 200$ and variable number $r = 500$, in order to study the performance of VBSSLQR and compare it with existing methods, including linear regression with a lasso penalty (Lasso), linear regression with an adaptive lasso penalty (ALasso), quantile regression with a lasso penalty (QRL), quantile regression with an adaptive lasso penalty (QRAL), Bayesian regularized quantile regression with a lasso penalty (BLQR), and Bayesian regularized quantile regression with an adaptive lasso penalty (BALQR). The data in the simulation studies were generated using Equation (1), in which the covariate vector x_i was randomly generated from the multivariate normal distribution $N(\mathbf{0}, \Sigma)$ with the (k, l) th of Σ being $0.5^{|k-l|}$. Among these covariates, we only considered the ten important explanatory variables that have significant impact on the dependent variable. We set the 1, 51, 101, 151, 201, 251, 301, 351, 401, and 451 predictors to be active, and their regression coefficients are $-3, -2.5, -2, -1.5, -1, 1, 1.5, 2, 2.5, 3$, and the rest are zero. In addition, we discuss the performance of various approaches in the case of two types of random error; namely, independent and identically distributed (i.i.d.) random errors and heterogeneous random errors.

3.1. Independent and Identically Distributed Random Errors Random

In this subsection, with reference to [19,23], we set the random errors ε_i 's in Equation (1) to be independently and identically distributed and consider the following five different distributions with τ quantile being zero:

- The error $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(-\mu, 1)$ with μ being the τ quantile of $N(0, 1)$, for $i = 1, \dots, n$;
- The error $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{Laplace}(-\mu, 1)$ with μ being the τ quantile of $\text{Laplace}(0, 1)$, where $\text{Laplace}(a, b)$ denotes the Laplace distribution with location parameter a and scale parameter b ;
- The error $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} 0.1N(-\mu_1, 9) + 0.9N(-\mu_2, 1)$ with μ_1 and μ_2 respectively being the τ quantile of $N(0, 9)$ and $N(0, 1)$;
- The error $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} 0.1\text{Laplace}(-\mu_1, 9) + 0.9\text{Laplace}(-\mu_2, 1)$ with μ_1 and μ_2 respectively being the τ quantile of $\text{Laplace}(0, 9)$ and $\text{Laplace}(0, 1)$;
- The error $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{Cauchy}(-\mu, 0.2)$ with μ being the τ quantile of $\text{Cauchy}(0, 0.2)$, where $\text{Cauchy}(a, b)$ denotes the Cauchy distribution with location parameter a and scale parameter b ;

For all of the above error distributions with any $\tau \in (0.3, 0.5, 0.7)$, we ran 500 replications for each method, and evaluated the performance using two criteria. The first criterion was the median of mean absolute deviations (MMAD), which quantifies the general distance between the estimated conditional quantile and the true conditional quantile. Specifically, the mean absolute deviation (MAD) in any replication is defined as $\frac{1}{n} \sum_{i=1}^n |x_i\beta - x_i\hat{\beta}_\tau|$, where $\hat{\beta}_\tau$ is the estimate of regression coefficient β given τ . The second criterion of the mean of true positives (TP) and false positives (FP) was selected for each method.

Table 1 shows the median and standard deviation (SD) of MADs estimated using each method for simulations with homogeneous errors. It is clear that our method was either optimal (bold) in all cases, especially in quantile $\tau = 0.3, 0.7$, or that our approach was significantly superior to the other six methods. When the error distribution was normal, a Laplace and normal mixture, the MMAD of the BALQR method was suboptimal. When the error distribution was Cauchy, the MMAD of the QRL approach was suboptimal, but Table 2 shows that this sacrificed the complexity of the model. When the error distribution was a Laplace mixture, the MMAD of the lasso approach was suboptimal; however, it selected an overfitting quantile regression model with about 30 FP variables, as can be seen from Table 2. It is particularly important to note that, in the case of high dimensional data, the MMAD for the quantile regression model with the lasso penalty or adaptive lasso penalty is not less than the MMAD for general linear models with a lasso penalty. Therefore, it is inappropriate to use a lasso penalty or adaptive lasso in this case.

Table 1. The median and standard deviation of 500 MADs estimated using various methods in simulations with i.i.d. errors.

Quantile	Error Distribution	Method						
		Lasso	ALasso	QRL	QRAL	BLQR	BALQR	VBSSLQR
$\tau = 0.3$	normal	0.60 (0.06)	0.89 (0.14)	0.88 (0.07)	3.90 (0.20)	1.28 (0.17)	0.55 (0.09)	0.21 (0.05)
	Laplace	0.72 (0.08)	0.97 (0.15)	1.06 (0.15)	3.95 (0.20)	1.44 (0.21)	0.68 (0.20)	0.24 (0.06)
	normal mixture	0.76 (0.09)	1.04 (0.14)	1.02 (0.15)	3.95 (0.21)	1.42 (0.19)	0.71 (0.16)	0.23 (0.05)
	Laplace mixture	0.89 (0.13)	1.13 (0.18)	1.24 (0.33)	4.06 (0.23)	1.69 (0.32)	1.07 (0.31)	0.26 (0.07)
	Cauchy	1.23 (4.59)	1.46 (12.07)	0.59 (0.62)	4.24 (12.40)	1.92 (1.07)	1.98 (4.95)	0.11 (0.60)
$\tau = 0.5$	normal	0.40 (0.05)	0.81 (0.16)	0.80 (0.07)	3.90 (0.21)	1.21 (0.19)	0.29 (0.07)	0.20 (0.05)
	Laplace	0.56 (0.08)	0.90 (0.17)	1.01 (0.16)	3.96 (0.21)	1.35 (0.22)	0.46 (0.23)	0.20 (0.05)
	normal mixture	0.53 (0.07)	0.89 (0.17)	0.96 (0.13)	3.95 (0.20)	1.36 (0.25)	0.41 (0.21)	0.21 (0.05)
	Laplace mixture	0.74 (0.13)	1.02 (0.20)	1.20 (0.22)	4.01 (0.23)	1.72 (0.32)	0.83 (0.32)	0.21 (0.06)
	Cauchy	1.31 (28.87)	1.44 (10.15)	0.48 (0.16)	4.40 (32.20)	2.026 (1.16)	4.28 (4.47)	0.07 (0.71)
$\tau = 0.7$	normal	0.61 (0.06)	0.93 (0.16)	0.88 (0.07)	3.88 (0.21)	1.22 (0.19)	0.56 (0.11)	0.20 (0.05)
	Laplace	0.71 (0.08)	1.02 (0.16)	1.06 (0.16)	3.97 (0.20)	1.36 (0.27)	0.67 (0.20)	0.25 (0.07)
	normal mixture	0.75 (0.09)	1.06 (0.17)	1.01 (0.14)	3.94 (0.21)	1.32 (0.23)	0.75 (0.19)	0.22 (0.05)
	Laplace mixture	0.89 (0.13)	1.20 (0.21)	1.24 (0.32)	4.01 (0.22)	1.65 (0.33)	0.99 (0.36)	0.26 (0.09)
	Cauchy	1.22 (8.78)	1.60 (8.43)	0.60 (0.60)	4.27 (13.65)	2.28 (1.25)	3.02 (2.91)	0.12 (1.14)

The bold represents the optimal result in each scenario.

In order to show the results of variable selection more intuitively, we introduced TP (true positives) and FP (false positives), to calculate the mean of TP and FP of 500 repeated simulations, respectively. Detailed results are shown in Table 2 below:

Table 2. Mean TP/FP of various methods for simulation with i.i.d. errors.

Quantile	Error Distribution	Method						
		Lasso TP/FP	ALasso TP/FP	QRL TP/FP	QRAL TP/FP	BLQR TP/FP	BALQR TP/FP	VBSSLQR TP/FP
$\tau = 0.3$	normal	10.00/38.24	9.80/0.20	10.00/175.01	5.00/95.33	8.48/0.00	10.00/0.02	10.00/0.02
	Laplace	10.00/40.43	9.75/0.34	9.99/169.28	5.00/94.88	8.06/0.00	9.86/0.13	10.00/0.01
	normal mixture	10.00/38.84	9.73/0.31	9.99/163.96	5.00/95.39	8.19/0.00	9.90/0.10	10.00/0.00
	Laplace mixture	10.00/37.99	9.62/0.71	9.87/146.31	5.00/95.25	7.43/0.00	9.50/0.70	10.00/0.01
	Cauchy	8.21/29.05	7.95/7.09	9.87/97.34	4.66/94.49	6.11/0.00	7.35/13.24	9.83/0.00
$\tau = 0.5$	normal	10.00/39.00	9.75/0.29	10.00/156.04	5.00/94.73	8.37/0.00	10.00/0.01	10.00/0.01
	Laplace	10.00/38.06	9.69/0.47	9.99/107.53	5.00/95.30	7.97/0.00	9.81/0.09	10.00/0.00
	normal mixture	10.00/39.03	9.69/0.57	10.00/112.74	5.00/95.33	8.10/0.00	9.86/0.06	10.00/0.00
	Laplace mixture	10.00/39.71	9.55/0.83	9.94/71.54	5.00/94.77	7.40/0.00	9.53/0.47	10.00/0.00
	Cauchy	8.19/29.31	8.04/7.06	10.00/8.50	4.55/95.67	5.53/0.00	6.03/22.94	9.77/0.00
$\tau = 0.7$	normal	10.00/38.85	9.70/0.41	10.00/174.87	5.00/95.40	8.41/0.00	9.99/0.03	10.00/0.01
	Laplace	10.00/37.68	9.64/0.67	9.99/168.15	5.00/95.05	7.91/0.00	9.77/0.13	10.00/0.01
	normal mixture	10.00/38.38	9.63/0.63	9.99/165.65	5.00/95.76	8.11/0.00	9.88/0.08	10.00/0.01
	Laplace mixture	10.00/39.64	9.41/1.22	9.88/150.33	5.00/95.56	7.32/0.00	9.50/1.03	9.99/0.01
	Cauchy	8.35/30.24	7.54/8.17	9.89/97.85	4.65/95.19	5.73/0.00	6.81/17.95	9.36/0.00

The bold represents the optimal result in each scenario.

It is possible to conclude from the results in Table 2 that the lasso method can generally select all true active variables, but it fits many false active variables; and when the random error is a Cauchy distribution, it cannot select all true active variables; A lasso approach can identify true active variables, but there are also some misjudged behaviors, especially when the random error distribution is a Cauchy distribution; Although the QRL approach can identify real active variables, it still contains many false active variables, and the model it identifies has a high complexity; the QRAL method cannot identify all true active variables, and incorrectly identifies some inactive variables; In the case of high-dimensional data, the

BLQR approach cannot select all the true active variables, but it also does not incorrectly select inactive variables. The true active variable selected using the BALQR method is better than using the BLQR method, but some inactive variables are incorrectly selected, especially when the random error distribution is the Cauchy distribution. Our VBSSLQR method not only had the smallest MMAD, but also could select true active variables and eliminate most false active variables. As a whole, our method was superior to the other six methods for variable selection, especially in quantile $\tau = 0.3, 0.7$, performing significantly better than BLQR, BALQR, and QRL.

3.2. Heterogeneous Random Errors

Now we consider the case of heterogeneous random errors, to demonstrate the performance of our method. In this subsection, the data were generated from the following model:

$$y_i = x_i^T \beta + (1 + u_i) \varepsilon_i, \tag{12}$$

where $u_i \stackrel{i.i.d.}{\sim} U(0, 1)$, in which $U(a, b)$ represents the uniform distribution with support set (a, b) . The design matrix x_i was generated in the same way as above, and the regression coefficient β was set as before. Furthermore, in the simulation study, we combined x_i and u_i ; that is, u_i was also a covariate. Finally, random error ε_i was also generated from the five different distributions defined in Section 3.1.

We also studied the performance of the quantile $\tau \in (0.3, 0.5, 0.7)$ under the different methods and simulated 500 times, to calculate the MMAD and mean of TP/FP. We list the experimental results in Tables 3 and 4.

Table 3. The median and standard deviation of 500 MADs were estimated via various methods for simulations with heterogeneous random errors.

Quantile	Error Distribution	Method						
		Lasso	ALasso	QRL	QRAL	BLQR	BALQR	VBSSLQR
$\tau = 0.3$	normal	0.91 (0.10)	1.13 (0.15)	1.31 (0.14)	3.96 (0.22)	1.57 (0.22)	0.90 (0.20)	0.30 (0.07)
	Laplace	1.08 (0.13)	1.24 (0.17)	1.61 (0.39)	4.09 (0.23)	1.93 (0.29)	1.23 (0.35)	0.36 (0.13)
	normal mixture	1.14 (0.14)	1.32 (0.17)	1.55 (0.35)	4.06 (0.24)	1.92 (0.31)	1.31 (0.35)	0.33 (0.11)
	Laplace mixture	1.36 (0.21)	1.51 (0.23)	2.34 (0.60)	4.27 (0.26)	2.30 (0.42)	2.05 (0.52)	0.39 (0.24)
	Cauchy	1.91 (102.51)	1.94 (33.45)	1.17 (0.80)	4.80 (19.32)	2.49 (1.11)	3.90 (3.39)	0.15 (0.98)
$\tau = 0.5$	normal	0.61 (0.08)	0.93 (0.17)	1.21 (0.12)	3.96/0.21	1.39 (0.27)	0.50 (0.22)	0.30 (0.07)
	Laplace	0.87 (0.13)	1.12 (0.20)	1.53 (0.25)	4.09 (0.23)	1.84 (0.33)	1.06 (0.35)	0.29 (0.08)
	normal mixture	0.81 (0.12)	1.08 (0.19)	1.44 (0.21)	4.05 (0.23)	1.77 (0.31)	1.08 (0.37)	0.31 (0.08)
	Laplace mixture	1.12 (0.21)	1.31 (0.25)	1.84 (0.32)	4.28 (0.29)	2.25 (0.41)	1.65 (0.56)	0.32 (0.12)
	Cauchy	1.93 (2.23)	1.93 (107.59)	0.74 (0.27)	4.93 (15.90)	2.88 (1.23)	3.72 (4.80)	0.09 (0.88)
$\tau = 0.7$	normal	0.92 (0.09)	1.20 (0.17)	1.32 (0.14)	3.97/0.20	1.38 (0.26)	0.86 (0.21)	0.31 (0.09)
	Laplace	1.07 (0.13)	1.31 (0.18)	1.61 (0.37)	4.09 (0.22)	1.95 (0.33)	1.29 (0.31)	0.37 (0.15)
	normal mixture	1.14 (0.14)	1.40 (0.19)	1.53 (0.33)	4.08 (0.23)	1.85 (0.36)	1.31 (0.29)	0.34 (0.10)
	Laplace mixture	1.38 (0.20)	1.58 (0.23)	2.16 (0.61)	4.32 (0.28)	2.37 (0.52)	1.86 (0.50)	0.40 (0.25)
	Cauchy	1.76 (3.76)	2.01 (17.31)	1.24 (0.76)	4.65 (31.22)	4.01 (1.30)	4.31 (18.21)	0.16 (1.04)

The bold represents the optimal result in each scenario.

For heterogeneous random errors, our approach was still the best, similarly to for i.i.d. random errors. It is noteworthy that the VBSSLQR method was more robust than the other methods, and our method had the smallest MMAD change compared to the case of i.i.d. random errors. We can see that the MMAD of our method remained basically unchanged, while the MMAD of the other six methods differed by more than 0.5 compared to the i.i.d. random errors in some states.

We also investigated the mean of TP and FP for heterogeneous random errors under different quantiles τ . The results are listed in Table 4, which shows that the effect of variable selection of heterogeneous random errors was slightly lower than the effect of variable

selection of i.i.d. random errors under the same sample size, but our method still provided the best selection results.

Table 4. Mean TP/FP of the various methods for simulations with heterogeneous random errors.

Quantile	Error Distribution	Method						
		Lasso TP/FP	ALasso TP/FP	QRL TP/FP	QRAL TP/FP	BLQR TP/FP	BALQR TP/FP	VBSSLQR TP/FP
$\tau = 0.3$	normal	10.00/40.76	9.72/0.35	9.98/170.29	5.00/95.23	7.97/0.00	9.75/0.13	10.00/0.03
	Laplace	10.00/38.80	9.54/0.83	9.73/140.09	4.99/94.59	7.23/0.00	9.31/1.20	9.97/0.05
	normal mixture	10.00/38.79	9.59/0.74	9.74/136.63	5.00/95.82	7.29/0.00	9.35/1.23	9.98/0.04
	Laplace mixture	9.95/37.96	9.17/2.07	8.82/76.98	4.96/95.42	6.53/0.00	8.51/4.48	9.83/0.05
	Cauchy	7.46/26.86	7.15/10.23	9.56/66.04	4.43/95.56	4.86/0.00	5.73/24.57	9.52/0.00
$\tau = 0.5$	normal	10.00/39.21	9.67/0.44	9.99/142.56	5.00/95.56	7.93/0.00	9.86/0.16	10.00/0.05
	Laplace	10.00/39.03	9.49/0.93	9.81/89.76	5.00/95.33	7.07/0.00	9.32/1.09	9.99/0.01
	normal mixture	10.00/38.44	9.51/0.86	9.83/85.73	4.99/95.27	7.41/0.00	9.28/0.96	9.99/0.02
	Laplace mixture	9.96/38.91	9.26/2.08	9.33/31.13	4.95/95.76	6.23/0.00	8.75/4.13	9.97/0.02
	Cauchy	7.24/25.47	7.23/9.37	9.99/4.30	4.26/95.95	4.54/0.00	5.92/24.88	9.60/0.00
$\tau = 0.7$	normal	10.00/40.58	9.59/0.52	9.98/171.43	5.00/95.53	7.88/0.00	9.76/0.16	10.00/0.07
	Laplace	10.00/39.20	9.44/1.14	9.74/144.66	4.99/95.20	6.87/0.00	9.26/0.97	9.95/0.06
	normal mixture	10.00/38.29	9.38/1.14	9.76/138.14	5.00/95.49	7.04/0.00	9.30/1.30	9.99/0.02
	Laplace mixture	9.96/38.88	9.08/2.37	8.93/85.78	4.96/95.02	5.95/0.00	8.69/3.50	9.81/0.05
	Cauchy	7.64/27.34	7.16/10.19	9.61/68.75	4.52/96.36	3.59/0.00	5.79/24.92	9.46/0.00

The bold represents the optimal result in each scenario.

We also calculated the mean execution times of various Bayesian quantile regressions under different quantile τ for different distributions of random errors, and list the results in Table A1 of Appendix C, which illustrates that our proposed VBSSLQR approach was a lot more efficient than BLQR and BALQR, for which we sampled MCMC 1000 times and discarded the first 500 times, and made a statistical inference based on 500 samples (the experimental study shows that the algorithm converged after 500 samples). In order to illustrate the feasibility of applying our proposed VBSSLQR to cases with smaller effect sizes, we changed the above active predictors to 1, while the other settings remained unchanged. The performance of our proposed method is shown in Table A2, which illustrates that the results were not significantly different from those of Tables 1 and 2 when random errors had an independent identical distribution, and slightly worse than those of Tables 3 and 4 when the random errors were heterogeneous.

4. Examples

In this section, we analyzed a real dataset containing information about crime in various cities across the United States. This dataset is accessible from the University of Irvine machine learning repository (<http://archive.ics.uci.edu/ml/datasets/communities+and+crime>, accessed on 1 May 2023). We calculated the per capita rate of violent crimes by dividing the total number of violent crimes by the population of each city. The violent crimes considered in our analysis were those classified as murder, rape, robbery, and assault, as per United States law. The observed individuals were communities. The dataset has 116 variables, where the first four columns are response, name of the community, code of county, and code of community; the middle features are demographic information about each community, such as population, age, race, and income; and the final columns are regions. According to the source: “the data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR”. This dataset has been applied for quantile regression [30]. The dataset is available at

- train set: <https://academics.hamilton.edu/mathematics/ckuruwit/Data/Crime/train.csv> (accessed on 1 May 2023).

- test set: <https://academics.hamilton.edu/mathematics/ckuruwit/Data/Crime/test.csv> (accessed on 1 May 2023).

Our dependent variable of interest was the murder rate of each city, denoted as y_i for the i th city. As the murder rate denotes the most dangerous violent crimes, we choose this variable. Studying factors correlated with the response dependent variable is of significant importance for the public and law enforcement agencies.

To adapt the data to our model, we preprocessed the data as follows:

- Delete columns from the data set that contain missing data.
- Delete the data when the response variable y_i equals 0, because this is not an issue of interest to us.
- Transform y_i : $y'_i = \log \frac{y_i}{1-y_i}$ and let y'_i be the new response variable.
- Convert some qualitative variables into quantitative variables.
- Standardized covariates.

After the above data preprocessing, we obtained 1060 observation objects and 95 covariates in the training set, and we obtained 122 observation objects and 95 covariates in the testing set. We implemented a quantile regression model between the 96 predictors (including intercept) and the response y'_i with different quantiles.

In this section, we compare QRL, QRAL, BLQR, and BALQR with our method VBSSLQR for real datasets, all with quantile penalty regression. Under different quantiles $\tau \in (0.1, 0.3, 0.5, 0.7, 0.9)$, we compared the performance of the different approaches. We counted the root mean squared error (RMSE) of each method under each quantile τ and the number of selected active variables, to evaluate the performance of each approach on the test set, where the RMSE was evaluated using $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{\tau i})^2}$ with $\hat{y}_{\tau i}$ being the fitted value of response y_i under quantile τ . Finally, the results are listed in Table 5 below.

Table 5. RMSE of fitting test dataset and the number of active variables for the various methods for the real dataset.

Method	RMSE/Number of Active Variables				
	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
QRL	1.36/64	0.94/65	1.09/66	0.93/57	1.21/53
QRAL	3.11/37	3.05/40	3.08/37	2.84/36	3.10/32
BLQR	1.34/1	1.08/1	1.03/1	1.08/1	1.30/1
BALQR	1.67/2	1.08/1	1.02/1	1.08/1	1.30/1
VBSSLQR	1.34/12	0.94/9	0.78/12	0.83/7	1.14/9

The bold represents the optimal result in each scenario.

To visually show the results listed in Table 4, the best results are in bold under each quantile τ . Clearly, our method performed better than the other methods for all quantiles, and the active variables selected by our method were suitable, which means that our method was very competitive compared with the other methods. The BLQR and BALQR approaches could only identify the intercept, and they could not identify the variables that really affected the response quantile. The QRL and QRAL methods were prone to overfitting, because they recognized too many active variables. Finally, the efficiency of our method with real data was also significantly higher than that of the other approaches. Although BLQR, QRL, and our proposed method had the same RMSE performance at $\tau = 0.1, 0.3$, BLQR showed underfitting and QRL showed overfitting. Thus, we believe there is sufficient evidence to show that our method was very competitive with the other approaches.

Similarly to [30], we list the active variables selected under each quantile in Table 6. Thus, Table 6 shows the variable selection of our proposed method with real data.

Table 6. Crime data analysis: variable selection.

Quantile Level (τ)	Quantile Specific Variables			
0.1	racepctblack	pctUrban	PctLess9thGrade	TotalPctDiv
	PctNotHSGrad	PctOccupMgmtProf	FemalePctDiv	PctPersDenseHous
	NumInShelters	PctHousOccup	PctBornSameState	
0.3	racepctblack	NumInShelters	FemalePctDiv	PctPersDenseHous
	TotalPctDiv	PctHousOwnOcc	RentLowQ	MedRent
0.5	racePctWhite	racePctHisp	FemalePctDiv	TotalPctDiv
	PctWorkMom	PctSpeakEnglOnly	HousVacant	PctVacantBoarded
	RentLowQ	MedRent	NumInShelters	
0.7	racePctWhite	racePctAsian	FemalePctDiv	TotalPctDiv
	PctVacantBoarded		NumInShelters	
0.9	racePctWhite	racePctAsian	indianPerCap	PctOccupManu
	MalePctDivorce	PctHousOccup	PctVacantBoarded	NumInShelters

In the above table, our method selected only a small number of predictors at each quantile level. Notably, only the variable “NumInShelters” had an impact on the response at all quantiles; the variables “FemalePctDiv” and “TotalPctDiv” had an impact on the response at $\tau = 0.1, 0.3, 0.5, 0.7$; the variable “PctVacantBoarded” had an impact on the response at $\tau = 0.5, 0.7, 0.9$; the variable “racePctWhite” had an impact on the response at $\tau = 0.5, 0.7, 0.9$; and the other variables affected a few quantiles of response. Therefore, the five variables selected from the quantile regression model were

- PctVacantBoarded: percentage of households that are vacant and boarded up to prevent vandalism.
- NumInShelters: number of shelters in the community.
- FemalePctDiv: percentage of females who are divorced.
- TotalPctDiv: percentage of people who are divorced.
- racePctWhite: percentage of people of white race.

In order to obtain a better understanding of these five common variables, we plot their correlation to the response.

Figure 1 depicts the relationship between the five common variables and the murder rate. Only the correlation between the NumInShelters and murder rate is not obvious, with MalePctDivorce, RentLowQ, racePctWhite, and PctVacantBoarded significantly affecting the murder rate. The variable racePctWhite and murder rate are negatively correlated, while FemalePctDiv, TotalPctDiv, PctVacantBoarded, and murder rate are positively correlated. This result was in accordance with the practical situation, and it can be seen that our results were basically consistent with that of [30]. Our method could more comprehensively select important variables under the same quantile. Therefore, the percentage of females who are divorced, percentage of persons who are divorced, the percentage of households that are vacant and boarded up to prevent from vandalism, and the percentage of people with white race affect the murder rate.

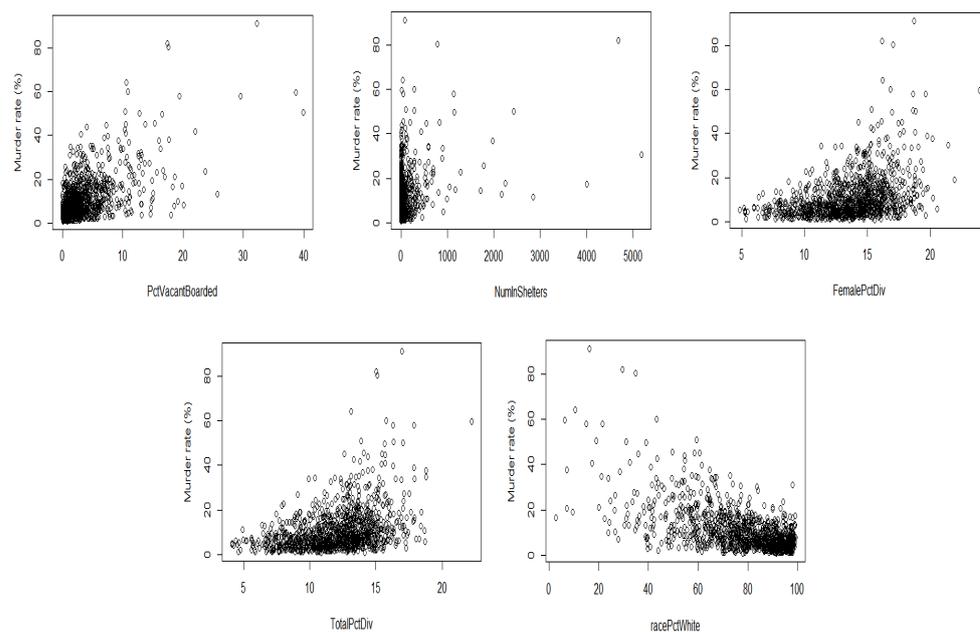


Figure 1. Correlation between murder rate and common variables.

5. Conclusions

In this paper, we propose variational Bayesian spike-and-slab lasso quantile regression variable selection and estimation. This method applies spike-and-slab lasso prior to each regression coefficient β , thus punishing the regression coefficient β . The spike prior distribution we choose is a small variance Laplace distribution, while the slab prior distribution is a large variance Laplace distribution [26]. It is precisely because it punishes each regression coefficient β that it has a powerful variable selection function, but this also brings a problem of low efficiency, especially when introducing a spike-and-slab prior into quantile regression (note, the algorithm efficiency of quantile regression is inherently lower than that of general regression approaches). In order to solve the problem of inefficiency (influenced by both the quantile regression and spike-and-slab lasso prior) and to make the algorithm feasible. We introduced a variational Bayesian method, to approximate the posterior distribution of each parameter. The simulation studies and real data analyses illustrated that the quantile regression with the spike-and-slab lasso penalty based on variational Bayesian method performed effectively and exhibited a robust competitiveness compared with other approaches (Bayesian method or non-Bayesian method, quantile regression, or nonquantile regression), especially in the case of high-dimensional data. In future research work, it would be significant to improve the interpretability and computational efficiency of ultra-high-dimensional quantile regression based on VB and dimension reduction techniques.

Author Contributions: Conceptualization, D.D. and A.T.; methodology, A.T. and D.D.; software, D.D. and J.Y.; validation, A.T., D.D. and J.Y.; formal analysis, D.D. and A.T.; investigation, D.D., J.Y. and A.T.; Preparation of the original work draft, A.T. and D.D.; visualization, D.D. and J.Y.; supervision, funding acquisition, D.D., A.T. and J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 11961079), Yunnan University Multidisciplinary Team Construction Projects in 2021, and Yunnan University Quality Improvement Plan for Graduate Course Textbook Construction (CZ22622202).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research data are available on the website <http://archive.ics.uci.edu/ml/datasets/communities+and+crime>, accessed on 1 May 2023.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Deduction

Based on the mean-field variational Formula (9), from the joint density (8), it is required to induce the following variational distributions.

$$\begin{aligned}
 q_1(\beta_j, \gamma_j) &\propto \exp \left\{ E_{-\beta_j, -\gamma_j} \left[\sum_{i=1}^n \log N(y_i | \mathbf{x}_i^\top \boldsymbol{\beta} + k_1 z_i, k_2 \sigma z_i) + \gamma_j \log N(\beta_j | 0, h_{0j}^2) \right. \right. \\
 &\quad \left. \left. + (1 - \gamma_j) \log N(\beta_j | 0, h_{0j}^2) + \log B(\gamma_j | 1, \pi_\gamma) \right] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2k_2} E_\sigma(\sigma^{-1}) \left[\sum_{i=1}^n E_{z_i}(z_i^{-1}) x_{ij}^2 \beta_j^2 + 2 \sum_{i=1}^n x_{ij} (E_{z_i}(z_i^{-1}) \right. \right. \\
 &\quad \left. \left. (y_i - \mathbf{x}_{i,(-j)}^\top \boldsymbol{\mu}_{\beta_{(-j)}}) - k_1) \beta_j \right] - \frac{\gamma_j}{2} E_{h_{0j}^2}(\log 2\pi h_{0j}^2) - \frac{\gamma_j}{2} E_{h_{0j}^2} \left(\frac{1}{h_{0j}^2} \right) \beta_j^2 \right. \\
 &\quad \left. - \frac{(1 - \gamma_j)}{2} E_{h_{1j}^2}(\log 2\pi h_{1j}^2) - \frac{(1 - \gamma_j)}{2} E_{h_{1j}^2} \left(\frac{1}{h_{1j}^2} \right) \beta_j^2 + \gamma_j E_{\pi_\gamma}(\log \pi_\gamma) \right. \\
 &\quad \left. + (1 - \gamma_j) E_{\pi_\gamma}(\log(1 - \pi_\gamma)) \right\} \\
 &\propto \exp \left\{ \left[-\frac{1}{2k_2} E_\sigma(\sigma^{-1}) \sum_{i=1}^n E_{z_i}(z_i^{-1}) x_{ij}^2 - \frac{\gamma_j}{2} E_{h_{0j}^2} \left(\frac{1}{h_{0j}^2} \right) - \frac{(1 - \gamma_j)}{2} E_{h_{1j}^2} \left(\frac{1}{h_{1j}^2} \right) \right] \beta_j^2 \right. \\
 &\quad \left. - \frac{1}{k_2} \sum_{i=1}^n x_{ij} (E_{z_i}(z_i^{-1}) (y_i - \mathbf{x}_{i,(-j)}^\top \boldsymbol{\mu}_{\beta_{(-j)}}) - k_1) \beta_j + \gamma_j E_{\pi_\gamma}(\log \pi_\gamma) \right. \\
 &\quad \left. + (1 - \gamma_j) E_{\pi_\gamma}(\log(1 - \pi_\gamma)) - \frac{\gamma_j}{2} E_{h_{0j}^2}(\log 2\pi h_{0j}^2) - \frac{(1 - \gamma_j)}{2} E_{h_{1j}^2}(\log 2\pi h_{1j}^2) \right\},
 \end{aligned}$$

where $\boldsymbol{\mu}_{\beta_{(-j)}} = E_{\beta_{(-j)}}(\boldsymbol{\beta}_{(-j)})$. therefore, given $\gamma_j = k$, for $k = 0$ and 1 , then

$$\begin{aligned}
 q_1(\beta_j | \gamma_j = k) &\propto \exp \left\{ E_{-\beta_j | \gamma_j = k} \left[\sum_{i=1}^n \log N(y_i | \mathbf{x}_i^\top \boldsymbol{\beta} + k_1 z_i, k_2 \sigma z_i) + k \log N(\beta_j | 0, h_{0j}^2) \right. \right. \\
 &\quad \left. \left. + (1 - k) \log N(\beta_j | 0, h_{0j}^2) \right] \right\},
 \end{aligned}$$

if $\gamma_j = 0$ then

$$\begin{aligned}
 q_1(\beta_j | \gamma_j = 0) &\propto \exp \left\{ \left[-\frac{1}{2k_2} E_\sigma(\sigma^{-1}) \sum_{i=1}^n E_{z_i}(z_i^{-1}) x_{ij}^2 - \frac{1}{2} E_{h_{1j}^2 | \gamma_j = 0} \left(\frac{1}{h_{1j}^2} \right) \right] \beta_j^2 \right. \\
 &\quad \left. - \frac{1}{k_2} \sum_{i=1}^n x_{ij} (E_{z_i}(z_i^{-1}) (y_i - \mathbf{x}_{i,(-j)}^\top \boldsymbol{\mu}_{\beta_{(-j)}}) - k_1) \beta_j \right\}, \\
 \beta_j | \gamma_j = 0 &\stackrel{\text{i.i.d.}}{\sim} N(\mu_{1j}, \sigma_{1j}^2),
 \end{aligned}$$

if $\gamma_j = 1$ then

$$q_1(\beta_j | \gamma_j = 1) \propto \exp \left\{ \left[-\frac{1}{2k_2} E_\sigma(\sigma^{-1}) \sum_{i=1}^n E_{z_i}(z_i^{-1}) x_{ij}^2 - \frac{1}{2} E_{h_{0j}^2 | \gamma_j = 1} \left(\frac{1}{h_{0j}^2} \right) \right] \beta_j^2 \right\}$$

$$\left. -\frac{1}{k_2} \sum_{i=1}^n x_{ij} (\mathbb{E}_{z_i}(z_i^{-1})(y_i - \mathbf{x}_{i,(-j)}^\top \boldsymbol{\mu}_{\beta_{(-j)}}) - k_1) \beta_j \right\},$$

$$\beta_j | \gamma_j = 1 \stackrel{\text{i.i.d.}}{\sim} N(\mu_{0j}, \sigma_{0j}^2),$$

therefore $\beta_j \stackrel{\text{i.i.d.}}{\sim} \sum_{k=0}^1 \mathbb{P}(\gamma_j = k) N(\mu_{kj}, \sigma_{kj}^2)$, for $j = 0, 1, \dots, r$, where

$$\mu_{kj} = \sigma_{kj}^2 \frac{\mathbb{E}_\sigma(\sigma^{-1})}{k_2} \sum_{i=1}^n \left[\mathbb{E}_{z_i}(z_i^{-1})(y_i - \mathbf{x}_{i,(-j)}^\top \boldsymbol{\mu}_{\beta_{(-j)}}) - k_1 \right] x_{ij},$$

$$\sigma_{kj}^2 = \left[\mathbb{E}_{h_{kj}^2 | \gamma_j = 1-k} (h_{kj}^{-2}) + \frac{\mathbb{E}_\sigma(\sigma^{-1})}{k_2} \sum_{i=1}^n x_{ij}^2 \mathbb{E}_{z_i}(z_i^{-1}) \right]^{-1},$$

for $k = 1$ and 2 . similarly, if $\gamma_j = 1$ then

$$q_2(h_{0j}^2 | \gamma_j = 1) \propto \exp \left\{ \mathbb{E}_{-h_{0j}^2 | \gamma_j = 1} \left[\log N(\beta_j | 0, h_{0j}^2) + \log \text{Exp}(h_{0j}^2 | \frac{\lambda_0^2}{2}) \right] \right\}$$

$$\propto (h_{0j}^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\mathbb{E}_{\beta_j | \gamma_j = 1}(\beta_j^2)}{2h_{0j}^2} - \frac{\mathbb{E}_{\lambda_0^2}(\lambda_0^2)}{2} h_{0j}^2 \right\}$$

$$= (h_{0j}^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\mu_{0j}^2 + \sigma_{0j}^2}{h_{0j}^2} + \mathbb{E}_{\lambda_0^2}(\lambda_0^2) h_{0j}^2 \right] \right\},$$

$$h_{0j}^2 | \gamma_j = 1 \stackrel{\text{i.i.d.}}{\sim} \text{GIG}(\frac{1}{2}, \mathbb{E}_{\lambda_0^2}(\lambda_0^2), \mu_{0j}^2 + \sigma_{0j}^2),$$

if $\gamma_j = 0$ then

$$q_3(h_{0j}^2 | \gamma_j = 0) \propto (h_{0j}^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbb{E}_{\lambda_0^2}(\lambda_0^2) h_{0j}^2 \right\},$$

$$h_{0j}^2 | \gamma_j = 0 \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\frac{1}{2} \mathbb{E}_{\lambda_0^2}(\lambda_0^2)),$$

therefore

$$q_2(h_{0j}^2) = \mathbb{P}(\gamma_j = 1) \text{GIG}(\frac{1}{2}, \mathbb{E}_{\lambda_0^2}(\lambda_0^2), \sigma_{0j}^2 + \mu_{0j}^2) + \mathbb{P}(\gamma_j = 0) \text{Exp}(\frac{1}{2} \mathbb{E}_{\lambda_0^2}(\lambda_0^2)).$$

similarly,

$$q_3(h_{1j}^2) = \mathbb{P}(\gamma_j = 0) \text{GIG}(\frac{1}{2}, \mathbb{E}_{\lambda_1^2}(\lambda_1^2), \sigma_{1j}^2 + \mu_{1j}^2) + \mathbb{P}(\gamma_j = 1) \text{Exp}(\frac{1}{2} \mathbb{E}_{\lambda_1^2}(\lambda_1^2)),$$

where μ_{kj} and σ_{kj}^2 are the same as above, for $k = 1$ and 0 , from $q(\beta_j, \gamma_j)$ then

$$q_1(\beta_j, \gamma_j = 1) = C \exp \left\{ -\frac{1}{2} \sigma_{0j}^{-2} \beta_j^2 + \frac{\mu_{0j}}{\sigma_{0j}^2} \beta_j + \mathbb{E}_{\pi_\gamma}(\log \pi_\gamma) - \frac{1}{2} \mathbb{E}_{h_{0j}^2 | \gamma_j = 1}(\log h_{0j}^2) \right\},$$

integral out β_j obtain

$$\mathbb{P}(\gamma_j = 1) = \int q(\beta_j, \gamma_j = 1) d\beta_j$$

$$= C \sqrt{\sigma_{0j}^2} \exp \left\{ \mu_{0j}^2 (2\sigma_{0j}^2)^{-1} + \mathbb{E}_{\pi_\gamma}(\log \pi_\gamma) - \frac{1}{2} \mathbb{E}_{h_{0j}^2 | \gamma_j = 1}(\log h_{0j}^2) \right\}.$$

similarly,

$$p(\gamma_j = 0) = \int q(\beta_j, \gamma_j = 0) d\beta_j \\ = C \sqrt{\sigma_{1j}^2} \exp \left\{ \mu_{1j}^2 (2\sigma_{1j}^2)^{-1} + E_{\pi_\gamma}(\log(1 - \pi_\gamma)) - \frac{1}{2} E_{h_{1j}^2 | \gamma_j = 0}(\log h_{1j}^2) \right\}.$$

letting

$$\zeta_j = \log p(\gamma_j = 0) - \log p(\gamma_j = 1) \\ = E_{\pi_\gamma}(\log(1 - \pi_\gamma)) - E_{\pi_\gamma}(\log \pi_\gamma) + \frac{1}{2} E_{h_{0j}^2 | \gamma_j = 1}(\log h_{0j}^2) \\ - \frac{1}{2} E_{h_{1j}^2 | \gamma_j = 0}(\log h_{1j}^2) + \frac{1}{2} (\log \sigma_{1j}^2 - \log \sigma_{0j}^2) + \frac{1}{2} (\mu_{1j}^2 (\sigma_{1j}^2)^{-1} - \mu_{0j}^2 (\sigma_{0j}^2)^{-1}),$$

then

$$p(\gamma_j = 1) = \frac{1}{1 + e^{\zeta_j}}, \quad p(\gamma_j = 0) = \frac{e^{\zeta_j}}{1 + e^{\zeta_j}},$$

therefore $\gamma_j \stackrel{i.i.d.}{\sim} B(1, (1 + e^{\zeta_j})^{-1})$, for $j = 0, 1, \dots, r$, where μ_{kj} and σ_{kj}^2 are the same as above, for $k = 1$ and 2 .

$$q_5(\lambda_0^2) \propto \exp \left\{ E_{-\lambda_0^2} \left[\sum_{j=0}^r \log \text{Exp}(h_{0j}^2 | \frac{\lambda_0^2}{2}) + \log \pi(\lambda_0^2) \right] \right\} \\ \propto (\lambda_0^2)^{\nu_{\lambda_0} + r + 1 - 1} \exp \left\{ - \frac{\sum_{j=0}^r E_{h_{0j}^2}(h_{0j}^2)}{2} \lambda_0^2 - \lambda_0^2 \right\} \\ \lambda_0^2 \sim \text{Ga}(\nu_{\lambda_0} + r + 1, \frac{\sum_{j=0}^r E_{h_{0j}^2}(h_{0j}^2)}{2} + 1),$$

similarly,

$$\lambda_1^2 \sim \text{Ga}(\nu_{\lambda_1} + r + 1, \frac{\sum_{j=0}^r E_{h_{1j}^2}(h_{1j}^2)}{2} + 1),$$

where ν_{λ_0} and ν_{λ_1} are hyperparameters.

$$q_7(\pi_\gamma) \propto \exp \left\{ E_{-\pi_\gamma} \left[\sum_{j=0}^r \log B(\gamma_j | 1, \pi_\gamma) + \log \pi(\pi_\gamma) \right] \right\} \\ = \exp \left\{ \sum_{j=0}^r E_{\gamma_j}(\gamma_j) \log \pi_\gamma + (r + 1 - \sum_{j=0}^r E_{\gamma_j}(\gamma_j)) \log(1 - \pi_\gamma) \right. \\ \left. + (a - 1) \log \pi_\gamma + (b - 1) \log(1 - \pi_\gamma) \right\} \\ = (\pi_\gamma)^{a - 1 + \sum_{j=0}^r E_{\gamma_j}(\gamma_j)} (1 - \pi_\gamma)^{r + b + 1 - 1 - \sum_{j=0}^r E_{\gamma_j}(\gamma_j)},$$

therefore $\pi_\gamma \sim \text{Be}(a + \sum_{j=0}^r E_{\gamma_j}(\gamma_j), r + b + 1 - \sum_{j=0}^r E_{\gamma_j}(\gamma_j))$ where a, b are hyperparameters.

$$q_4(z_i) \propto \exp \left\{ E_{-z_i} \left[\log N(y_i | \mathbf{x}_i^\top \boldsymbol{\beta} + k_1 z_i, k_2 \sigma z_i) + \log E(z_i | \sigma^{-1}) \right] \right\} \\ \propto z_i^{\frac{1}{2} - 1} \exp \left\{ - \frac{1}{2} \left[\frac{E_\sigma(\sigma^{-1})}{k_2} ((k_1^2 + 2k_2) z_i + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\mu}_\beta)^2 + \mathbf{x}_i^\top \boldsymbol{\Sigma}_\beta \mathbf{x}_i}{z_i}) \right] \right\},$$

therefore $z_i \stackrel{\text{i.i.d.}}{\sim} \text{GIG}(\frac{1}{2}, a_{z_i}, b_{z_i})$, for $i = 1, \dots, n$, where

$$a_{z_i} = \frac{E_{\sigma}(\sigma^{-1})}{k_2} (k_1^2 + 2k_2), b_{z_i} = \frac{E_{\sigma}(\sigma^{-1})}{k_2} [(y_i - \mathbf{x}_i^T \boldsymbol{\mu}_{\beta})^2 + \mathbf{x}_i^T \boldsymbol{\Sigma}_{\beta} \mathbf{x}_i],$$

with k_1 and k_2 being constants and $\boldsymbol{\Sigma}_{\beta}$ is a $(p + 1) \times (p + 1)$ diagonal matrix with j th entry being $\text{Var}(\beta_j)$.

$$\begin{aligned} q_8(\sigma) &\propto \exp \left\{ E_{-\sigma} \left[\sum_{i=1}^n \log N(y_i | \mathbf{x}_i^T \boldsymbol{\beta} + k_1 z_i, k_2 \sigma z_i) + \log E(z_i | \sigma^{-1}) + \log \pi(\sigma) \right] \right\} \\ &\propto \left(\frac{1}{\sigma} \right)^{\frac{n}{2} + n + a_{\sigma} - 1} \exp \left\{ - \frac{1}{2k_2} \left[k_1^2 E_{z_i}(z_i) + E_{z_i}(z_i^{-1}) ((y_i - \mathbf{x}_i^T \boldsymbol{\mu}_{\beta})^2 + \mathbf{x}_i^T \boldsymbol{\Sigma}_{\beta} \mathbf{x}_i) \right. \right. \\ &\quad \left. \left. - 2k_1 (y_i - \mathbf{x}_i^T \boldsymbol{\mu}_{\beta}) \right] \sigma^{-1} - \sum_{i=1}^n E_{z_i}(z_i) \sigma^{-1} - b_{\sigma} \sigma^{-1} \right\}, \end{aligned}$$

therefore $\sigma \sim \text{IG}(\frac{3}{2}n + a_{\sigma}, c_{\sigma})$, where a_{σ}, b_{σ} are hyperparameters and

$$\begin{aligned} c_{\sigma} &= \frac{1}{2k_2} \sum_{i=1}^n \{ k_1^2 E_{z_i}(z_i) + E_{z_i}(z_i^{-1}) [(y_i - \mathbf{x}_i^T \boldsymbol{\mu}_{\beta})^2 + \mathbf{x}_i^T \boldsymbol{\Sigma}_{\beta} \mathbf{x}_i] \\ &\quad - 2k_1 (y_i - \mathbf{x}_i^T \boldsymbol{\mu}_{\beta}) \} + \sum_{i=1}^n E_{z_i}(z_i) + b_{\sigma}. \end{aligned}$$

Appendix B. Expectation

The expectation of some parameter functions about variational posteriors:

Since $\beta_j \stackrel{\text{i.i.d.}}{\sim} \sum_{k=0}^1 p(\gamma_j = k) N(\mu_{kj}, \sigma_{kj}^2)$, therefore

$$E_{\beta_j}(\beta_j) = \sum_{k=0}^1 p(\gamma_j = k) E_{\beta_j | \gamma_j = k}(\beta_j) = \sum_{k=0}^1 p(\gamma_j = k) \mu_{kj},$$

$$\text{Var}_{\beta_j}(\beta_j) = p(\gamma_j = 1) \sigma_{0j}^2 + p(\gamma_j = 0) \sigma_{1j}^2 + p(\gamma_j = 1) p(\gamma_j = 0) (\mu_{0j} - \mu_{1j})^2,$$

$$E_{\beta_j | \gamma_j = k}(\beta_j^2) = \mu_{kj}^2 + \sigma_{kj}^2, k = 0, 1.$$

Since $h_{0j}^2 \stackrel{\text{i.i.d.}}{\sim} p(\gamma_j = 1) \text{GIG}(\frac{1}{2}, E_{\lambda_0^2}(\lambda_0^2), \sigma_{0j}^2 + \mu_{0j}^2) + p(\gamma_j = 0) \text{Exp}(\frac{1}{2} E_{\lambda_0^2}(\lambda_0^2))$, therefore

$$E_{h_{0j}^2} \left(\frac{1}{h_{0j}^2} \right) = \sum_{k=0}^1 p(\gamma_j = k) E_{h_{0j}^2 | \gamma_j = k} \left(\frac{1}{h_{0j}^2} \right),$$

where $E_{h_{0j}^2 | \gamma_j = 1} \left(\frac{1}{h_{0j}^2} \right) = \sqrt{E_{\lambda_0^2}(\lambda_0^2) (E_{\beta_j | \gamma_j = 1}(\beta_j^2))^{-1}}$ and $E_{h_{0j}^2 | \gamma_j = 0} \left(\frac{1}{h_{0j}^2} \right) = 2[E_{\lambda_0^2}(\lambda_0^2)]^{-1}$,

$$E_{h_{0j}^2 | \gamma_j = 1}(h_{0j}^2) = \sqrt{E_{\beta_j | \gamma_j = 1}(\beta_j^2) (E_{\lambda_0^2}(\lambda_0^2))^{-1}} + (E_{\lambda_0^2}(\lambda_0^2))^{-1}.$$

Since $h_{1j}^2 \stackrel{\text{i.i.d.}}{\sim} p(\gamma_j = 0) \text{GIG}(\frac{1}{2}, E_{\lambda_1^2}(\lambda_1^2), \sigma_{1j}^2 + \mu_{1j}^2) + p(\gamma_j = 1) \text{Exp}(\frac{1}{2} E_{\lambda_1^2}(\lambda_1^2))$, therefore

$$E_{h_{1j}^2} \left(\frac{1}{h_{1j}^2} \right) = \sum_{k=0}^1 p(\gamma_j = k) E_{h_{1j}^2 | \gamma_j = k} \left(\frac{1}{h_{1j}^2} \right),$$

where $E_{h_{1j}^2 | \gamma_j = 1} \left(\frac{1}{h_{1j}^2} \right) = 2[E_{\lambda_1^2}(\lambda_1^2)]^{-1}$ and $E_{h_{1j}^2 | \gamma_j = 0} \left(\frac{1}{h_{1j}^2} \right) = \sqrt{E_{\lambda_1^2}(\lambda_1^2) (E_{\beta_j | \gamma_j = 0}(\beta_j^2))^{-1}}$,

$$E_{h_{1j}^2 | \gamma_j = 0}(h_{1j}^2) = \sqrt{E_{\beta_j | \gamma_j = 0}(\beta_j^2) (E_{\lambda_1^2}(\lambda_1^2))^{-1}} + (E_{\lambda_1^2}(\lambda_1^2))^{-1}.$$

Since $\sigma \sim \text{IG}(\frac{3}{2}n + a_\sigma, c_\sigma)$, therefore

$$E_\sigma(\sigma^{-1}) = (\frac{3}{2}n + a_\sigma) \frac{1}{c_\sigma}, E_\sigma(\sigma) = (\frac{3}{2}n + a_\sigma - 1)^{-1} c_\sigma,$$

Since $z_i \stackrel{\text{i.i.d.}}{\sim} \text{GIG}(\frac{1}{2}, a_{z_i}, b_{z_i})$, therefore

$$E_{z_i}(z_i^{-1}) = \sqrt{a_{z_i}(b_{z_i})^{-1}}, E_{z_i}(z_i) = \sqrt{b_{z_i}(a_{z_i})^{-1}} + (a_{z_i})^{-1}.$$

Since $\lambda_k^2 \sim \text{Ga}(v_{\lambda_k} + r + 1, \frac{\sum_{j=0}^r E_{h_{kj}^2}(h_{kj}^2)}{2} + 1)$, for $k = 0$ and 1 , therefore

$$E_{\lambda_k^2}(\lambda_k^2) = (r + 1 + v_{\lambda_k})(1 + \frac{1}{2} \sum_{j=0}^r E_{h_{kj}^2}(h_{kj}^2))^{-1}.$$

Since $\pi_\gamma \sim \text{Be}(a + \sum_{j=0}^r E_{\gamma_j}(\gamma_j), r + b + 1 - \sum_{j=0}^r E_{\gamma_j}(\gamma_j))$, therefore

$$E_{\pi_\gamma}(\log \pi_\gamma) = \Psi(a + \sum_{j=0}^r E_{\gamma_j}(\gamma_j)) - \Psi(a + b + r + 1),$$

$$E_{\pi_\gamma}(\log(1 - \pi_\gamma)) = \Psi(r + b + 1 - \sum_{j=0}^r E_{\gamma_j}(\gamma_j)) - \Psi(a + b + r + 1),$$

$$E_{\pi_\gamma}(\pi_\gamma) = (a + \sum_{j=0}^r E_{\gamma_j}(\gamma_j))(a + r + b + 1)^{-1}.$$

Since $\gamma_j \stackrel{\text{i.i.d.}}{\sim} B(1, (1 + e^{\zeta_j})^{-1})$, therefore

$$E_{\gamma_j}(\gamma_j) = p(\gamma_j = 1) = (1 + e^{\zeta_j})^{-1}.$$

Appendix C. Efficiency Comparison between Bayesian Quantile Regression Methods

Table A1. Mean execution times of the various Bayesian quantile regression methods in the simulation.

Quantile	Error Distribution	Method					
		i.i.d. Random Errors			Heterogenous Random Errors		
		BLQR	BALQR	VBSSLQR	BLQR	BALQR	VBSSLQR
$\tau = 0.3$	normal	146.45 s	144.56 s	18.06 s	146.48 s	174.88 s	28.77 s
	Laplace	145.26 s	144.62 s	18.82 s	145.86 s	173.44 s	30.08 s
	normal mixture	144.52 s	145.27 s	18.11 s	143.48 s	147.63 s	28.99 s
	Laplace mixture	145.86 s	146.28 s	19.07 s	145.82 s	171.69 s	30.38 s
	Cauchy	145.24 s	144.63 s	16.06 s	145.89 s	174.68 s	25.64 s
$\tau = 0.5$	normal	144.62 s	144.52 s	15.30 s	144.67 s	174.39 s	25.23 s
	Laplace	144.54 s	145.26 s	15.00 s	144.57 s	172.84 s	24.51 s
	normal mixture	144.62 s	145.26 s	15.39 s	144.63 s	173.62 s	25.21 s
	Laplace mixture	147.68 s	145.83 s	15.14 s	147.62 s	172.25 s	25.13 s
	Cauchy	145.21 s	144.53 s	12.38 s	145.86 s	173.42 s	20.72 s
$\tau = 0.7$	normal	147.29 s	144.37 s	18.74 s	146.48 s	173.46 s	31.05 s
	Laplace	144.62 s	145.27 s	19.89 s	144.64 s	174.61 s	33.86 s
	normal mixture	145.84 s	145.24 s	19.00 s	145.27 s	172.84 s	31.60 s
	Laplace mixture	147.57 s	147.98 s	19.88 s	147.68 s	173.42 s	34.31 s
	Cauchy	145.87 s	142.86 s	16.87 s	146.28 s	173.45 s	28.12 s

The bold represents the optimal result in each scenario and “s” denotes seconds.

Appendix D. Simulation Studies for Cases with Smaller Effect Sizes

Table A2. The performance of our proposed method for cases with smaller effect sizes.

Quantile	Distribution	Errors					
		i.i.d. Random Errors			Heterogenous Random Errors		
		MMAD(sd)	TP	FP	MMAD(sd)	TP	FP
$\tau = 0.3$	normal	0.21(0.05)	10.00	0.13	0.33(0.10)	10.00	0.25
	Laplace	0.25(0.08)	10.00	0.15	0.42(0.34)	9.56	0.37
	normal mixture	0.23(0.06)	10.00	0.07	0.36(0.16)	9.93	0.23
	Laplace mixture	0.29(0.10)	9.99	0.16	0.98(0.50)	8.32	0.38
	Cauchy	0.12(0.55)	9.43	0.00	0.16(0.76)	8.73	0.01
$\tau = 0.5$	normal	0.20(0.06)	10.00	0.16	0.32(0.09)	10.00	0.29
	Laplace	0.20(0.06)	10.00	0.06	0.32(0.11)	9.99	0.20
	normal mixture	0.22(0.06)	10.00	0.10	0.34(0.10)	10.00	0.21
	Laplace mixture	0.22(0.07)	10.00	0.06	0.33(0.26)	9.76	0.17
	Cauchy	0.07(0.50)	9.55	0.00	0.10(0.69)	9.00	0.00
$\tau = 0.7$	normal	0.21(0.06)	10.00	0.13	0.34(0.10)	10.00	0.37
	Laplace	0.25(0.08)	10.00	0.15	0.40(0.18)	9.91	0.38
	normal mixture	0.23(0.06)	10.00	0.16	0.36(0.12)	9.99	0.31
	Laplace mixture	0.27(0.09)	10.00	0.17	0.43(0.30)	9.66	0.33
	Cauchy	0.11(0.61)	9.38	0.00	0.15(0.67)	9.13	0.00

References

- Koenker, R.; Bassett, G. Regression quantile. *Econometrica* **1978**, *46*, 33–50. [\[CrossRef\]](#)
- Buchinsky, M. Changes in the united-states wage structure 1963–1987—Application of quantile regression. *Econometrica* **1994**, *62*, 405–458. [\[CrossRef\]](#)
- Thisted, R.; Osborne, M.; Portnoy, S.; Koenker, R. The gaussian hare and the laplacian tortoise: Computability of squared-error versus absolute-error estimators—Comments and rejoinders. *Stat. Sci.* **1997**, *12*, 296–300.
- Koenker, R.; Hallock, K. Quantile regression. *J. Econ. Perspect.* **2001**, *15*, 143–156.
- Yu, K.; Moyeed, R. Bayesian quantile regression. *Stat. Probab. Lett.* **2001**, *54*, 437–447. [\[CrossRef\]](#)
- Taddy, M.A.; Kottas, A. A bayesian nonparametric approach to inference for quantile regression. *J. Bus. Econ. Stat.* **2010**, *28*, 357–369. [\[CrossRef\]](#)
- Hu, Y.; Gramacy, R.B.; Lian, H. Bayesian quantile regression for single-index models. *Stat. Comput.* **2013**, *23*, 437–454. [\[CrossRef\]](#)
- Lee, E.R.; Noh, H.; Park, B.U. Model selection via bayesian information criterion for quantile regression models. *J. Am. Stat. Assoc.* **2014**, *109*, 216–229. [\[CrossRef\]](#)
- Frank, I.; Friedman, J. A statistical view of some chemometrics regression tools. *Technometrics* **1993**, *35*, 109–135. [\[CrossRef\]](#)
- Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [\[CrossRef\]](#)
- Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360.
- Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser.-Stat. Methodol.* **2005**, *67*, 301–320. [\[CrossRef\]](#)
- Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [\[CrossRef\]](#)
- Koenker, R. Quantile regression for longitudinal data. *J. Multivar. Anal.* **2004**, *91*, 74–89. [\[CrossRef\]](#)
- Wu, Y.; Liu, Y. Variable selection in quantile regression. *Stat. Sin.* **2009**, *19*, 801–817.
- Park, T.; Casella, G. The bayesian lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [\[CrossRef\]](#)
- Leng, C.; Tran, M.N.; Nott, D. Bayesian adaptive Lasso. *Ann. Inst. Stat. Math.* **2014**, *66*, 221–244. [\[CrossRef\]](#)
- Li, Q.; Xi, R.; Lin, N. Bayesian regularized quantile regression. *Bayesian Anal.* **2010**, *5*, 533–556. [\[CrossRef\]](#)
- Alhamzawi, R.; Yu, K.; Benoit, D.F. Bayesian adaptive Lasso quantile regression. *Stat. Model.* **2012**, *12*, 279–297. [\[CrossRef\]](#)
- Ishwaran, H.; Rao, J. Spike and slab variable selection: Frequentist and bayesian strategies. *Ann. Stat.* **2005**, *33*, 730–773. [\[CrossRef\]](#)
- Ray, K.; Szabo, B. Variational bayes for high-dimensional linear regression with sparse priors. *J. Am. Stat. Assoc.* **2022**, *117*, 1270–1281. [\[CrossRef\]](#)
- Yi, J.; Tang, N. Variational bayesian inference in high-dimensional linear mixed models. *Mathematics* **2022**, *10*, 463. [\[CrossRef\]](#)
- Xi, R.; Li, Y.; Hu, Y. Bayesian quantile regression based on the empirical likelihood with spike and slab priors. *Bayesian Anal.* **2016**, *11*, 821–855. [\[CrossRef\]](#)
- Koenker, R.; Machado, J. Goodness of fit and related inference processes for quantile regression. *J. Am. Stat. Assoc.* **1999**, *94*, 1296–1310. [\[CrossRef\]](#)

25. Tsionas, E. Bayesian quantile inference. *J. Stat. Comput. Simul.* **2003**, *73*, 659–674. [[CrossRef](#)]
26. Rockova, V. Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *Ann. Stat.* **2018**, *46*, 401–437. [[CrossRef](#)]
27. Alhamzawi, R.; Ali, H.T.M. The bayesian adaptive lasso regression. *Math. Biosci.* **2018**, *303*, 75–82. [[CrossRef](#)]
28. Parisi, G.; Shankar, R. Statistical field theory. *Phys. Today* **1988**, *41*, 110. [[CrossRef](#)]
29. Beal, M.J. Variational Algorithms for Approximate Bayesian Inference. Ph.D Thesis, University of London, London, UK, 2003.
30. Kuruwita, C.N. Variable selection in the single-index quantile regression model with high-dimensional covariates. *Commun.-Stat.-Simul. Comput.* **2021**, *52*, 1120–1132. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.