1. Introduction

Quantile regression has received increasing attention from both theoretical and empirical points of view. It is a statistical procedure intended to estimate conditional quantile of a response distribution, given a set of covariates, and to conduct inference about them.

In regression, when we deal with highly skewed data distribution, traditional mean regression may not explore interesting aspects of the relationship between the response variable and the available covariates. Furthermore, in the presence of outliers, the evaluation of the response average becomes much more complicated. However, since a set of quantiles often provides a complete description of the response distribution conditionally to the given covariates, quantile regression offers a practical alternative to the traditional mean regression.

Quantile regression was initially introduced by Koenker and Bassett [1], where inference proceeds by minimizing a check loss function. Based on their theory, many other approaches have been emerged to study the regression quantiles in both frequentist and Bayesian points of view.

In the frequentist framework, different techniques have been developed in the literature to infer regression quantiles. To estimate linear regression quantiles, Koenker and Bassett [1] used linear programming technique, whereas Gutenbrunner and Jurecková [2] considered the corresponding dual problem whose solution represents the quantile coefficients estimator. From theoretical point of view, asymptotic normality and consistency properties are proven in Koenker and Bassett [1], Zhou et al. [3], and Koenker and Xiao [4].

For nonparametric quantile regression, spline methods are commonly used for quantile regression fitting (see Nychka et al. [5], Koenker and Portnoy [6], and Koenker et al. [7]). Oh et al. [8] proposed, in univariate and bivariate settings, an approximation of the nonlinear optimization problem by a sequence of penalized least squares type nonparametric mean estimation problems. Koenker and Mizera [9] explored the estimation of triograms in bivariate setting. Koenker et al. [10] proposed an additive models estimation method for quantile regression based on selecting the smoothing parameter and constructing confidence bands for nonparametric components, in univariate and bivariate settings. The univariate regression quantile spline is extended to problems with several covariates in He and Ng [11] and Li et al. [12] where, in this latter, regression quantiles are considered in reproducing kernel Hilbert spaces (RKHS).

In the Bayesian framework, different methods have been successfully emerged to study quantile regression. This was pioneered by Yu and Moyeed [13], where the authors specified
the Bayesian quantile regression model and assumed that the error terms are iid according to the asymmetric Laplace distribution (ALD), denoted by $\mathcal{ALD}(\mu, \sigma, p)$ with $\mu \in \mathbb{R}$ being its $p$–th quantile, $\sigma$ is the scale parameter, and $p \in [0,1]$ are the asymmetry parameters. They also showed that the Bayesian approach yields a proper full conditional distribution, even for an improper uniform prior on the quantile coefficients and they used the Metropolis-Hastings algorithm for implementation. Further, Kozumi and Kobayashi [14] and Bernardi et al. [15] considered that the scale parameter $\sigma$ of the ALD is unknown. They used the location-scale mixture representation of the ALD, which enables developing a Gibbs sampling algorithm for the Bayesian quantile regression model. In the literature, we can also find other methods that extended quantile regression model. For censored data, Kobayashi et al. [16] proposed Tobit quantile regression with endogenous variables and Li and Bradic [17] developed regression adjustment for local adaptive representation of random forests. In continuous dependent variable case and for variable selection and coefficient estimation, Alhamzawi et al. [18] used adaptive LASSO quantile regression and, recently, Sayed-Ahmed [19] applied this approach for small sample sizes, whereas Abbas and Thaher [20] developed Bayesian adaptive Tobit regression. Benoit and Van den Poel [21] proposed Bayesian quantile regression methods for binary response data and Alhamzawi and Ali [22] adapted the quantile regression model to deal with longitudinal ordinal data.

All these methods mentioned above, from both frequentist and Bayesian points of view, studied single regression quantile. However, there are many applications where tackling several quantiles at the same time is needed. For this direction, if one uses methods that deal with single quantile regression to estimate regression quantiles separately, results may not be satisfactory. Indeed, it is known that the separate estimation of a set of quantiles may break the quantile monotonicity property (see He [23]), which means that quantile curves may cross.

Fortunately, many approaches have been proposed to overcome the crossing problem. In the frequentist framework, Liu and Wu [24] and Bondell et al. [25] considered monotonicity constraints for a fixed number of quantiles in both linear and nonparametric cases. Sangnier et al. [26] proposed a nonparametric method that estimates multiple conditional quantiles simultaneously when assuming that the quantile functions span an RKHS. In this work, they choose an appropriate matrix-valued kernel taking into account the distance between quantiles.

On the other hand, in simultaneous Bayesian framework, Reich et al. [27] proposed a two-stage semiparametric method for linear quantile regression, which is one of the first methods addressing the crossing problem of quantiles. At the first stage, they used Koenker and Bassett’s [1] method to estimate quantile coefficients. At the second stage, these coefficients are reestimated with Bernstein polynomial basis function with constraints on their coefficients. Later, Reich [28] proposed a fully Bayesian approach based on interpolating the quantile function on different quantile levels using a piecewise Gaussian basis function. Tokdar et al. [29] proposed a simultaneous linear quantile regression method for a univariate explanatory variable with a bounded and convex domain. They showed that a linear quantile, $q_p(x) = \beta_0(\tau) + x\beta(\tau)$, is monotonically increasing if and only if $\beta_0(\tau)$ and $\beta(\tau)$ are linear combination of two monotonic increasing functions in $\tau \in [0,1]$. Afterwards, Yang and Tokdar [30] extended this method to the multivariate explanatory case. Another two-stage method is proposed by Rodrigues and Fan [31]. At the first stage, they used the standard Bayesian quantile regression of Yu and Moyeed [13] fitted separately at the different quantile levels. They introduced induced quantiles, at the second stage, using a relation between quantiles of the ALD distribution. At the same stage, they used the Gaussian process regression that monotonizes quantile functions by borrowing strength from the nearby ones. However, this method is still affected by the outputs of the first stage since different likelihoods are used for initial estimates. For nonparametric quantile regression, a cubic B-spline method is considered by Muggeo et al. [32], solving an $L_2$ penalisation problem with monotonicity constraint on spline coefficients. Many recent works on simultaneous Bayesian quantile regression can be found in literature such as Das and Ghosal [33], Das and Ghosal [34], Rodrigues et al. [35], and Petrella and Raponi [36].

In this paper, our contribution aims at developing a fully Bayesian approach that estimates multiple quantiles simultaneously in one step. Unlike the other Bayesian methods, our proposed method is not based on misspecification, i.e., the ALD is the real underlying response distribution. We used a relation between quantiles of the ALD distribution and Metropolis-Hastings within Gibbs algorithm for implementation. Our proposed method provides parallel quantile functions estimators and, thus, the noncrossing is a natural result. Our numerical results show, using specific empirical criteria, that the estimated regression quantiles respect the quantile level ordering.

The paper is organised as follows. Section 2 describes the method presenting the model and establishing the relation between quantiles of the ALD distribution used in the estimation procedure. The algorithm associated with the proposed method is given in Section 3, where we introduce the location-scale mixture representation of the ALD. In Section 4, simulated examples are provided to show the performance and the flexibility of the method, and the final section presents our conclusion.

2. Simultaneous Bayesian Estimation of Quantiles

2.1. Model. We consider the quantile regression model

$$Y_i = q_p(X_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

with $(\epsilon_i) \sim \mathcal{ALD}(\mu = 0, \sigma, p)$, where $\mu$, $\sigma$, and $p$, respectively, are the location, the scale, and the asymmetry
parameters. This leads to assuming that the response variable \( Y_i \), given the covariable \( X_i \), follows an ALD distribution:

\[
Y_i \mid (q_p(X_i), \sigma, p) \sim \mathcal{ALD}(q_p(X_i), \sigma, p)
\]

\[
\forall i = 1, \ldots, n,
\]

whose probability density function (p.d.f.) is given by

\[
f(y; q_p(X_i), \sigma, p) = \frac{p (1 - p)}{\sigma} \exp \left\{ -\frac{\rho_p(y - q_p(X_i))}{\sigma} \right\},
\]

where \( \rho_p(u) = u(p - 1)_{u < 0} \) is the check loss function. The explanatory variables \( X_i \)'s are supposed to be iid distributed according to a continuous distribution, \( P_X \), whose support is \( \mathcal{X}^d \), \( \mathcal{X} \subset \mathbb{R}, d \geq 1 \). In what follows, the function \( q_p(.) \) and the parameters \( \sigma, p, \) and \( P_X \) are supposed to be unknown.

In all subsequent sections, any quantity in bold represents a vector; capital letters denote random variables or vectors, whereas lowercase letters denote observed values of the corresponding random vectors.

### 2.2. Bayesian Procedure

#### 2.2.1. Likelihood.

To infer simultaneously \( s \) distinct quantiles with a fully Bayesian approach, say \( q_{r_1}, \ldots, q_{r_s} \), with distinct orders \( r_1, \ldots, r_s \), of the conditional distribution of \( Y \mid X \), one needs to characterize the likelihood through all these \( s \) unknown quantiles.

As explained below, this is done first by partitioning the whole sample into \( s \) subsamples, which requires a sufficient number of observations, second by using the relation between any two quantiles of the ALD distribution, and third by rewriting the whole likelihood through \( s \) terms, where the \( j \)-th term only depends on \( q_{r_j} \).

1. Consider the following partition of the whole sample through \( s \) subsamples:

\[
(X_{I_j}, Y_{I_j}) = \{(X_i, Y_i)_{i \in I_j}\}, \quad \forall j \in \{1, \ldots, s\},
\]

where \( I_j = \{(j - 1)r + 1, \ldots, jr\} \) with \( I_j \cap I_k = \emptyset \), \( \forall j \neq k, \) and \( j, k \in \{1, \ldots, s\} \); suppose that \( r = n/s \) is integer, without loss of generality.

2. To characterize the likelihood through all quantiles of interest, we use the relation that links any \( r \)-th quantile to the \( p \)-th quantile of the underlying ALD distribution (see Yu and Zhang [37] and Rodrigues and Fan [31]):

\[
q_r(X) = q_p(X) + \sigma g(\tau, p),
\]

where \( g(\tau, p) = (1/(1 - p)) \log(\tau/p)1_{0 < \tau < p} - (1/p) \log((1 - \tau)/(1 - p))1_{p < \tau < 1} \).

From (5), we rewrite the model given by (1) as follows:

\[
Y_{I_j} = q_{r_{j_i}} - \sigma \beta_{r_j} + \epsilon_{I_j}, \quad \forall j = 1, \ldots, s,
\]

(6)

where

\[
Y_{I_j} = \left(Y_{I_j}\right)_{i \in I_j},
\]

\[
q_{r_{j_i}} = \left(q_{r_{j_i}}(X_i)\right)_{i \in I_j},
\]

\[
\beta_{r_j} = \left(\beta_{r_j}\right)_{i \in I_j},
\]

\[
\epsilon_{I_j} = \left(\epsilon_{I_j}\right)_{i \in I_j}.
\]

Then, the likelihood associated with the model given by (1) is rewritten as the product of \( s \) likelihoods; each one only depends on one subsample:

\[
L(\{q_{r_{j_i}}, \ldots, q_{r_{j_k}}\}, \sigma, p; x_1, \ldots, x_s, y_1, \ldots, y_s)
\]

\[
= \prod_{j=1}^{s} \prod_{i \in I_j} \frac{p (1 - p)}{\sigma} \exp \left\{ -\frac{\rho_p(y_i - q_{r_{j_i}}(x_i) - \sigma g(\tau_j, p))}{\sigma} \right\}.
\]

(8)

#### 2.2.2. Priors.

In this section, we specify the prior distributions for all unknown quantities. It is worth mentioning that \( P_X \) is unknown even if it is not our primary interest, and therefore it should require a prior; but since any prior on \( P_X \) that is independent of the prior on \( (q_{r_{j_i}}, \sigma, p) \) would disappear upon marginalization of the posterior of \( (P_X, \{q_{r_{j_i}}\}_{j=1,\ldots,s}, \sigma, p) \) relatively to \( P_X \), we drop it in the sequel. Thus, it suffices to choose a prior distribution for \( q_{r_{j_i}} \), for \( j = 1, \ldots, s, \sigma, \) and \( p \).

1. For the quantiles, we distinguish between the linear and nonlinear cases.

   a. Linear case. \( q_{r_j}(X) = \beta_{r_j}^{(0)} + X^T \beta_{r_j}, j \in \{1, \ldots, s\} \), where \( X^T \) denotes the transpose of \( X \) and \( \beta_{r_j} \in \mathbb{R}^d \).

Due to (5), one should note that distinct quantiles differ only by the intercept \( \beta_{r_j}^{(0)} \), so that all quantiles of interest have the same slope \( (\beta_{r_1} = \ldots = \beta_{r_s} = \beta_{r_s}^{(0)}) \); then we consider the unknown vector parameter \( \hat{\beta} = (\beta_{r_1}^{(0)}, \ldots, \beta_{r_s}^{(0)}, \beta_{r_s}^{(0)}) \) of dimension \( s + d \) with Gaussian prior:

\[
\hat{\beta} \sim \mathcal{N}_s + d \left( \hat{\Sigma}_0 \right),
\]

(9)

with \( \hat{\Sigma}_0 \) being a positive definite square matrix of dimension \( s + d \).
(b) **Nonlinear case.** We define the function \( h \) on \([0, 1] \times \mathbb{R}^d \), by \( h(\tau, x) = q_\tau(x) \). We put a Gaussian process prior on \( h, h \sim \mathcal{GP}(0, k) \), with a zero-mean function and a covariance function \( k: [0, 1]^2 \times \mathbb{R}^d \rightarrow \mathbb{R} ; \) following Sangnier et al. [26], we choose \( k \) to be decomposable and its form is given by

\[
k((\tau, x), (\tau', x')) = k_x(x, x') \exp \left(-c(\tau - \tau')^2\right),
\]

where \( k_x(x, x') = \exp(-b\|x - x'\|^2) \), \( b \) and \( c \) are positive hyperparameters, and \( \| \cdot \| \) is the Euclidean norm on \( \mathbb{R}^d \). Here, \( k((\tau, x), (\tau', x')) \), \( i \neq j \), encodes the relation between the conditional quantiles \( q_\tau(x) \) and \( q_{\tau'}(x') \). As explained by Sangnier et al. [26], if \( c \rightarrow 0 \), the quantile curves are parallel so they do not cross. If \( c \rightarrow +\infty \), the quantiles are learned independently and then they may cross. Thus, the choice of \( c \) is important to control the crossing.

(2) \( \sigma \sim \mathcal{IG}(a_0, b_0) \) and \( a_0 > 0, b_0 > 0 \), where \( \mathcal{IG} \) denotes the inverse Gamma distribution with positive hyperparameters \( a_0 \) and \( b_0 \).

(3) \( p \sim \text{Beta}(\alpha_0, \lambda_0) \) and \( \alpha_0, \lambda_0 > 0 \).

### 3. Computations

To compute the full conditional distributions, we shall make use of the location-scale mixture of the ALD distribution (see Kozumi and Kobayashi [14]).

Let \( \omega \) be an exponential latent variable with parameter \( 1/\sigma \), denoted by \( \mathcal{E}(1/\sigma) \), and let \( Z \) be a standard normal variable such that \( \omega \) and \( Z \) are independent. If \( \epsilon \) has an ALD distribution, \( \epsilon \sim \mathcal{ALD}(0, \sigma, p) \); then, it can be represented as a mixture of normal variable:

\[
\epsilon = g_p \omega + \delta_p \sqrt{\omega \epsilon},
\]

where \( g_p = (1 - 2p)/p(1 - p) \) and \( \delta_p^2 = 2/\eta(1 - p) \).

Exploiting this augmented data structure, the model defined by the system of equations given by (6) admits, conditionally on \( \omega_i \),

\[
Y_i \sim \mathcal{ALD}( \omega_i, \sigma, p, \epsilon_i),
\]

the following Gaussian representation:

\[
Y_i = q_{\tau_i} + (1 - 2p)/p(1 - p) \omega_i + \delta_p \sqrt{\omega_i \epsilon_i},
\]

where \( \omega_i \) is a mixture of normal variables.

#### 3.1. Full Conditional Distributions

From the equations given in (12) and the priors defined in Section 2.2.2, we follow Bernardi et al. [15] to derive the full conditional distributions of \( \omega_n \) and \( (q_{\tau_i})_{j=1,\ldots,s} \). Yet, the full conditionals of \( \sigma \) and \( p \) are not tractable and this is carried out by including a Metropolis-Hasting step to the Gibbs sampler algorithm.

(1) For all \( j = 1, \ldots, s \), as well as \( i \in I_j \), set \( \omega_i = \omega_i^{-1} \), and considering the distribution \( \omega_i \sim \mathcal{E}(1/\sigma) \) as a prior on \( \omega_i \),

\[
v_i \sim \mathcal{FGau}(\Psi_i, \phi_p)
\]

with

\[
\Psi_i = \sqrt{\frac{\gamma_p^2 + 2\delta_p^2}{\phi_p^2 + 2\delta_p^2}},
\]

\[
\phi_p = \frac{\gamma_p^2 + 2\delta_p^2}{\delta_p^2 \sigma^2},
\]

where \( \mathcal{FGau} \) stands for the Inverse Gaussian distribution with \( \Psi_i > 0 \) and \( \phi_p > 0 \) as location and shape parameters.

(2) The conjugate Gaussian prior on quantiles provides Gaussian full conditional distributions in both linear and nonlinear cases.

(a) **Linear case.** Let us introduce some notations:

(i) \( Y_n = (Y_j)_{1 \leq j \leq n} \)

(ii) \( X = (X_{ij})_{1 \leq i \leq n, 1 \leq j \leq s} \) is the \( n \times (s + d) \) design matrix defined by

\[
X_{ij} = \begin{cases} \mathbb{I}_{kl} & \text{if } l = 1, \ldots, s, \\ \mathbb{I}_{kl} \mathbb{I}_{ls} & \text{if } l = s + 1, \ldots, s + d, \end{cases}
\]

(iii) \( \mathbf{e}_n = (g(\tau_j, p)\mathbb{I}_{kl})_{1 \leq j \leq s} \).

This allows to rewrite the system of equations (12) into the following format:

\[
Y_n = \mathbf{X} \beta - \sigma \mathbf{e}_n + \gamma_p \omega_n + \mathbf{e}_n^*,
\]

where \( \epsilon_n^* \sim \mathcal{N}_s(0, \Sigma^*) \) and where \( \Sigma^* = \text{diag}(\delta_p^2 \sigma \omega_n) \). With the zero-mean Gaussian prior distribution, the full conditional distribution on \( \beta \) is then Gaussian:

\[
\hat{\beta} \sim \mathcal{N}(\mathbf{X}^\top \mathbf{X}^{-1} \mathbf{X}^\top (Y_n + \sigma \mathbf{e}_n - \gamma_p \omega_n)),
\]

with

\[
\sigma^2_{\hat{\beta}} = (\mathbf{X}^\top \Sigma^* \mathbf{X} + \tilde{\Sigma}^*)^{-1}.
\]

(b) **Nonlinear case.** We shall use other extra notations:

\[
h_n = (h_{\tau_j})_{j=1,\ldots,s} = (h(\tau_j, x_j))_{1 \leq j \leq s},
\]

\[
k_n(\tau, x) = \left( k(\tau_j, x_j)_{1 \leq j \leq s} \right)_j=1,\ldots,s.
\]
This allows to rewrite the system of equations (12) into the following vector format:
\[ Y_n = h_n - \sigma g_n + \gamma_p \omega_n + e_n. \]  \hspace{1cm} (20)

Combining (20) with the Gaussian process prior on \( h \) leads to the following joint distribution of \((Y_n, h_n)\) conditional on \((x, \omega_n, \sigma, p)\):
\[
\begin{pmatrix}
    Y_n \\
    h_n
\end{pmatrix}
~ \mathcal{N}_{2n}
\begin{pmatrix}
    -\sigma g_n + \gamma_p \omega_n \\
    0_n
\end{pmatrix},
\end{equation}
\[ k_n (\tau, x) + \Sigma^* \begin{pmatrix}
    k_n (\tau, x) \\
    k_n (\tau, x)
\end{pmatrix}, \]  \hspace{1cm} (21)

Finally, classical calculations lead to the desired full conditional distribution:
\[
h_n \mid (x, y_n, \omega_n, \sigma, p) \sim \mathcal{N} \left( \tilde{\mu}_h, \tilde{\Sigma}_h \right), \]  \hspace{1cm} (22)

where
\[ \tilde{\mu}_h = k_n (\tau, x) (k_n (\tau, x) + \Sigma^*)^{-1} (y_n - \gamma_p \omega_n + \sigma g_n), \]
\[ \tilde{\Sigma}_h = k_n (\tau, x) - k_n (\tau, x) (k_n (\tau, x) + \Sigma^*)^{-1} k_n (\tau, x). \]  \hspace{1cm} (23)

The full conditional distribution, \( \pi_\sigma \) of \( \sigma \mid x, \omega_n, \sigma, p \), is proportional to
\[
\frac{1}{\sqrt{2\pi}^{n+1}} \exp \left\{ \frac{1}{2} \left( Y_n - q_{r,n} + \sigma g_n - \gamma_p \omega_n \right)^T \cdot \left( \Sigma^* \right)^{-1} \left( Y_n - q_{r,n} + \sigma g_n - \gamma_p \omega_n \right) \right\} \]  \hspace{1cm} (24)

where \( q_{r,n} = (q_{r,j})_{j=1,...,s} \).

The full conditional, \( \pi_p \) of \( p \mid x, \omega_n, \sigma \), is proportional to
\[
p^\alpha (1-p)^{\beta_0} \exp \left\{ \frac{1}{2} \left( Y_n - q_{r,n} + \sigma g_n - \gamma_p \omega_n \right)^T \cdot \left( \Sigma^* \right)^{-1} \left( Y_n - q_{r,n} + \sigma g_n - \gamma_p \omega_n \right) \right\}\]  \hspace{1cm} (25)

3.2. Algorithm. Due to (24) and (25), it is not possible to generate \( \sigma \) and \( p \) directly from their full conditional distribution. Therefore, they are simulated by incorporating a random walk Metropolis-Hastings step within Gibbs, as described in Algorithm 1.

For a subset \( E \subset \mathbb{R} \), denote by \( \mathcal{N}_E (\cdot, \cdot) \) the truncated version on \( E \) of the corresponding Gaussian distribution. Note that the choice of the truncated Gaussian distribution is classical.

It is worth to mention that the values of the scale parameters \( sd_\sigma \) and \( sd_\gamma \) are calibrated to achieve the equilibrium of the random walk Metropolis-Hastings step quickly; in fact, they are chosen neither too small nor too large so that the acceptance rate becomes practically stable.

The conditional posterior ratios of \( \sigma \) and \( p \) are given by
\[
P_\sigma \quad \text{and} \quad P_p
\]
\[
\begin{align*}
P_\sigma &= \frac{N_\sigma \left( \sigma_{\text{prop}} \mid y_n, x_n \left( q_{r,j} \right)_{j=1,...,s}, \omega_n^{(t+1)}, p^{(t)} \right)}{N_\sigma \left( \sigma^{(t)} \mid y_n, x_n \left( q_{r,j} \right)_{j=1,...,s}, \omega_n^{(t+1)}, p^{(t)} \right)} \\
P_p &= \frac{N_p \left( p_{\text{prop}} \mid y_n, x_n \left( q_{r,j} \right)_{j=1,...,s}, \omega_n^{(t+1)}, p^{(t)} \right)}{N_p \left( p^{(t)} \mid y_n, x_n \left( q_{r,j} \right)_{j=1,...,s}, \omega_n^{(t+1)}, p^{(t)} \right)}
\end{align*} \hspace{1cm} (26)

where \( N_\sigma (\cdot \mid \cdot) \) and \( N_p (\cdot \mid \cdot) \), are respectively, given up to a constant in (24) and (25), and the transition probabilities are given by
\[
\begin{align*}
Q_\sigma &= \frac{f_{y_n \mid (\sigma^{(t+1)}, \omega_n)}}{f_{y_n \mid (\sigma^{(t)}, \omega_n)}} \frac{1 - \phi (-\sigma_{\text{prop}} / sd_\sigma)}{1 - \phi (-\sigma^{(t)} / sd_\sigma)} \\
Q_p &= \frac{f_{y_n \mid (p_{\text{prop}}, \omega_n)}}{f_{y_n \mid (p^{(t)}, \omega_n)}} \frac{\phi \left( (1 - p_{\text{prop}}) / sd_p \right) - \phi \left( -p_{\text{prop}} / sd_p \right)}{\phi \left( (1 - p^{(t)}) / sd_p \right) - \phi \left( -p^{(t)} / sd_p \right)}
\end{align*} \hspace{1cm} (27)
where \( f_{\theta}(\cdot) \) denotes the pdf of \( N_{p}(\cdot, \cdot) \) and \( \phi(\cdot) \) denotes the cumulative distribution function of the standard normal distribution.

### 4. Simulation Study

In this section, we study the performance of our method in both linear and nonlinear quantile regressions. For the model given by (1), we shall consider three different designs for the \( p \)-th quantile:

1. **Univariate linear quantile:** \( q_p(X) = 1 + 2X \), with \( X \sim \mathcal{U}_{[-1,1]} \).
2. **Multivariate linear quantile:** \( q_p(X) = 1 + X^{T} \beta_p \), with \( \beta_p \in \mathbb{R}^d \), \( d = 10 \), and either \( X_l \sim \mathcal{U}_{[-1,1]} \) or \( X_l \sim \mathcal{N}(0, 1) \), \( \forall l \in \{1, \ldots, d\} \).
3. **Nonlinear quantile:** for \( X \sim \mathcal{U}_{[0,1]} \),
   \[
   q_p(X) = \cos\left(\frac{\sqrt{2} \pi \exp\left\{\frac{3}{2} X\right\}}{2}\right) \\
   + \left[\frac{1}{4} \exp\{2(X - 0.5)\} - \exp\{-1\}\left(\frac{1}{4} + \frac{X}{2}\right)\right]I_{(X < 0.5)} \\
   - \left(\frac{1}{4} + \frac{X}{2}\right) \exp\{-1\} - \left(\frac{1}{2} - X\right) \right]I_{(X \geq 0.5)}. \tag{28}
   
   For all designs, we generate independently 300 observations issued from the model defined in (1). All runs of Metropolis-Hastings within Gibbs algorithm consist in 20000 iterations, one-third of which are burn-in. The prior hyperparameters are chosen carefully to illustrate the intended objectives in all designs. We use the RMISE to evaluate the estimation performance. These other methods are performed using available codes in R Core Team (2017): qfunc available in quantreg package Koenker [39] for “K&B”, bayesQR function in bayesQR package Benoit et al. [40] for “Y&M”, and greg function in Bsquare package Smith et al. [41] for “R&S”.

For design 3, we compare our “SBQR” method with both the nonparametric quantile regression method of Muggeo et al. [32], denoted by “M&ST”, and the simultaneous noncrossing method of Rodrigues and Fan [31], denoted by “F&R”. We have implemented “M&ST” with quantregGrowth R package (see Muggeo [42]) and “F&R” with an own made code in R. We use features of the RMISE criterion to show how well “SBQR” performs among the other considered methods.

We note that the number of observations, \( n = 300 \), is chosen carefully to illustrate the intended objectives in all designs.

#### 4.1. Design 1: Univariate Linear Quantile Regression

The iid sample \( e_1, \ldots, e_n \) is generated according to \( \mathcal{A}(\mu = 0, \sigma = 0.1, p = 0.25) \). We propose to infer three quantiles that are close, namely, the quantiles of order \( \tau = 0.2, 0.3 \) respectively.

We fix \( sd_{\sigma} = sd_{\rho} = 0.1 \). To evaluate the convergence of our algorithm, we use three different seeds and parameter starting values to run three different chains and calculate \( \hat{R} \) of Gelman convergence diagnostic. Besides, we use other convergence diagnostics such as the autocorrelation analysis and the posterior plots.

As shown in the top panel of Figures 1, 2, and 3, all posterior distributions shrink at the true parameters value. Furthermore, in the middle panel of Figures 1, 2, and 3, the decrease of the empirical autocorrelation of posterior samples proves that the underlying chains are stationary. The bottom panels of Figures 1, 2, and 3 show that \( \hat{R} \) goes to 1 through the iterations, which confirms the convergence of the algorithm.

#### 4.2. Design 2: Multivariate Linear Quantile Regression

The second design is dedicated to the multivariate linear case; hence, we consider the multivariate distribution of posterior samples to prove that the underlying chains are stationary. We propose to infer two close quantiles of order \( \tau_1 = 0.1 \) and \( \tau_2 = 0.12 \) and to study the crossing throughout the following four methods: “K&B”, “Y&M”, “R&S”, and “SBQR”.

To achieve the desired posterior distribution through MCMC methods, we perform Y&M and R&S with different number of iterations: 1000 for Y&M and 10000 for R&S. For “R&S”, we use the logistic base distribution with 4 basis functions. For “SBQR”, we fix \( sd_{\sigma} = sd_{\rho} = 0.01 \) and \( \hat{S}_{\sigma} \) to be the identity matrix.
To compare the methods, we make use of the crossing loss criterion (see Sangnier et al. [26]) that measures how far $\hat{q}_{0.12}(X)$ goes below $\hat{q}_{0.10}(X)$:

$$\text{crossing loss} = \frac{1}{n} \sum_{i=1}^{n} \max (0, \hat{q}_{0.10}(X_i) - \hat{q}_{0.12}(X_i)).$$

(29)

For a given approach, the less the crossing loss is, the better is the method. As shown in Table 1, the crossing loss, by “K&B”, is significantly of 0.43%, which corresponds to 34 data points of $\hat{q}_{0.1}(X)$ which are above $\hat{q}_{0.1}(X)$. This percentage is considerably weakened when applying the separate Bayesian method “Y&M” (0.38%) but still has crossing quantiles (26 data points of $\hat{q}_{0.1}(X)$ are above $\hat{q}_{0.12}(X)$). However, for simultaneous estimation methods, as our proposed “SBQR” or “R&S” methods, the crossing loss becomes zero; this means that the simultaneous quantile estimation has the potential to make quantile crossing vanish. Moreover, the estimated quantiles are in the right order. While simultaneous approaches control the monotonicity property of quantiles in a certain sense, separate approaches do not, and they provide more than 8% of violation according to results previously discussed (11.33% of violation by “K&B” and 8.66% of violation by “Y&M”).

To go further, we compute the RMISE for each quantile order and for all methods. The “SBQR” has roughly the smallest RMISE in both quantile levels $\tau = 0.1$ and $\tau = 0.12$. The fact is that “R&S” may oversmooth when estimating the quantiles simultaneously and here, for this linear case, the oversmoothness impacts the flexibility of the method (high RMISE).
To see if the covariable support has an impact on results, we consider another simulation set in which $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}_d, \mathcal{I}_d)$, that is, support($\mathbf{X}$) = $\mathbb{R}^d$. Table 2 shows that “SBQR” behaves like in the previous example: a zero crossing loss and the smallest RMISE among all the methods.

We also consider another case with a different pair of quantile orders: $\tau_1 = 0.7$ and $\tau_2 = 0.8$. We turn back to the support $[-1,1]^d$ for $\mathbf{X}$. Table 3 shows similar results as the ones obtained for $\tau_1 = 0.10$ and $\tau_2 = 0.12$. Thus, “SBQR” still has the best behavior among the other methods in terms of crossing loss and RMISE.

It should be noted that the estimation of $\sigma$ and $p$ by “SBQR” is quite good, since their estimated values are near the true ones in the different treated cases.

### 4.3. Design 3: Nonparametric Quantile Regression

Considering the third design with $\mathbf{X} \sim \mathcal{U}_{[0,1]}$ and $\mathbf{e} \sim \mathcal{ALD}(\mu = 0, \sigma = 0.05, p = 0.75)$, we are interested in estimating quantile functions for orders $\tau = 0.10, 0.12, 0.15$, and 0.20.
We fix $sd_p = 0.05$ and $sd_{\sigma} = 0.005$ and we compare our "SBQR" method with two others: the "M&ST" method with three-order cubic B-splines and the "F&R" approach with smoothness parameter value equal to 0.1.

For each value of $\tau$, we evaluate the performance of these methods through the RMISE criterion. Table 4 shows that the RMISE values are significantly smaller for "SBQR" than the ones for "M&ST" and "F&R". It is worth noting, therefore,
that our method is significantly better than the two others in quantiles estimation; besides, it provides a reasonable estimation of $\sigma (\hat{\sigma} = 0.0535)$ and $\rho (\hat{\rho} = 0.7744)$.

Figure 4 gives the quantile curves estimators. While our “SBQR” method (left panel) provides quantile curves estimates that are close to the true ones (red-dashed lines), the 0.1-th quantile curve estimate given by “F&R” method (middle panel) is fairly distant from the true curve, especially when $x \in [0.2, 0.6]$. The same happens for “M&ST” method when $x \in [0.7, 1]$. However, there is no scarred crossing by any of these three methods, since they are tackling simultaneous quantile estimation techniques.

The issue that makes “F&R” method less flexible in simultaneous quantiles fitting is that, in the second stage, the final estimators are quite affected by the first stage output that may be badly estimated. For “M&ST”, the estimators are constructed iteratively, when solving the minimization problem, by adding constraints so that each subsequent quantile function does not cross with the previous one; this may cause overestimation of quantile curves. Our “SBQR” method does not suffer from these mentioned disadvantages, which explains its significant flexibility.

5. Conclusion

Our proposed estimation procedure, “SBQR”, for simultaneous Bayesian quantile regression guarantees the fundamental property of noncrossing. Assuming that the ALD is the underlying data distribution, this method enables characterizing the likelihood function by all quantiles of interest using the relation between two distinct quantiles. Using the location-scale normal mixture representation of the ALD distribution, we develop a Metropolis-Hastings within Gibbs algorithm to implement our method. Our simulation studies show good results that reflect the good performance of the method and the convergence of the algorithm. Against the crossing problem of estimated quantiles, our method has good performance compared with single quantile estimation methods like Koenker and Bassett [1] and Yu and Moyeed [13] methods. From the RMISE point of view, “SBQR” is very competitive in both linear and nonlinear quantile regression cases.

As future perspectives, we envisage extending this work to deal with any conditional distribution of $Y \mid X$.

Data Availability

The simulated data used to support the findings of this study are provided by R software and are included within the attached document.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The author would like to thank the Ph.D. supervisor Professor Ghislaine Gayraud, University of Technology of Compiègne,
France, for her support and comments that greatly assisted this paper.

**Supplementary Materials**

The simulated data used to support the findings of this study are provided by R software and are included within the attached document. *(Supplementary Materials)*

**References**


Submit your manuscripts at
www.hindawi.com