Bayesian Quantile Regression for Big Data Analysis

Yuanqi Chu, Xueping Hu, and Keming Yu

Abstract Quantile regression, which estimates various conditional quantiles of a response variable, including the median (0.5th quantile), is particularly useful when the conditional distribution is asymmetric or heterogeneous or fat-tailed or truncated. Bayesian methods for the inference of quantile regression have been receiving increasing attention from both theoretical and empirical viewpoints but facing the challenge of scaling up the existing methods when the data are too large to be processed by a single machine under many big data environments nowadays. In this paper, we explore Bayesian quantile regression (BQR) analysis via normal-inverse-gamma (NIG) distribution type of likelihood function, prior distribution and posterior distribution. We further develop the details of methods of BQR for massive data applications. The performance of proposed methods is evaluated via real data illustrations.

Keywords Quantile regression (QR) · Bayesian inference · Big data · Normal-inverse-gamma (*NIG*)

1 Introduction

Quantile regression (QR) estimates various conditional quantiles of a response or dependent random variable, including the median (0.5th quantile). Putting different quantile regressions together provides a more complete description of the underlying conditional distribution of the response than a simple mean regression. This is

Y. Chu · K. Yu (✉)
Department of Mathematics, Brunel University London, Middlesex UB8 3PH, UK
e-mail: keming.yu@brunel.ac.uk

Y. Chu
e-mail: yuanqi.chu@brunel.ac.uk

X. Hu · K. Yu
College of Mathematics and Physics, Anqing Normal University, Anqing 246133, People's Republic of China
e-mail: hxprob@163.com

R. Argiento et al. (eds.), *New Frontiers in Bayesian Statistics*, Springer Proceedings in Mathematics & Statistics 405, https://doi.org/10.1007/978-3-031-16427-9_2

particularly useful when the conditional distribution is asymmetric or heterogeneous or fat-tailed or truncated. Quantile regression has been widely used in statistics and numerous application areas ([3, 5, 11, 25] and among others). In the "big data" era for statistical science, the rich of data sources with many complicated data structures and the increase of extreme values and heterogeneity may see quantile regression methods more relevant than mean regression to dig deep into the data and grab information from it. In particular, with advanced power of computer, complicated quantile regression-based models could be developed under a Bayesian framework, and Bayesian quantile regression (BQR) has received increasing attention from both theoretical and empirical viewpoints with wide applications and variants (see [4, 10, 12, 17, 19, 23] and among others). So far, in the context of quantile regression, several methods have been developed for big data analysis ([6, 9, 22, 27] and among others), but little attention has been paid to such methodology under Bayesian inference paradigm.

In this paper, we propose a new approach of BQR for big data. This approach has its posterior distribution on the whole data as a joint posterior from M sub data split from the whole data. Section 2 introduces the likelihood function for BQR based on the location-scale mixture of normals for asymmetric Laplace distribution [15, 18]. Section 3 gives details of the normal-inverse-gamma (NIG) expressions of the prior and posterior distributions for BQR via informative g-prior [28]. Section 4 derives the posterior distribution on the whole data as a joint multiplication of the posterior obtained from M sub data split from the whole data via NIG summation operator, and provides big data based algorithms for BQR. Section 5 demonstrates the proposed approaches and algorithms via real data illustrations. Some concluding remarks are presented in Sect. 6.

2 Quantile Regression and Its Likelihood Function

Let $y_i, i = 1, \ldots, n$ be a continuous response variable and $\mathbf{x_i}$ a $k \times 1$ vector of predictors for the ith observation. The linear quantile regression model for the pth quantile can be denoted as $y_i = \mathbf{x_i}^T \boldsymbol{\beta} + \varepsilon_i$, where $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters of interest, and ε_i is the error term whose distribution is assumed to have zero pth quantile. The estimation for $\boldsymbol{\beta}$ is solved by minimizing $\sum_{i=1}^{n} \rho_p(y_i - \mathbf{x_i}^T \boldsymbol{\beta})$, where $\rho_p(u) = u\{p - I(u < 0)\}$ is the check function and $I(\cdot)$ denotes the indicator function. According to [24, 26], such minimization is equivalent to maximizing a likelihood function that is based on the asymmetric Laplace distribution (ALD) at specific value of p. Assume that errors $\varepsilon_i, i = 1, \ldots, n$ are $ALD(0, \sigma, p)$, with the likelihood given by

$$f(\boldsymbol{\varepsilon}|\sigma) \propto \sigma^{-n} \exp\{-\sum_{i=1}^{n} \frac{|\varepsilon_i| + (2p-1)\varepsilon_i}{2\sigma}\},$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$. Following [15, 18], we can represent ε_i as a location-scale mixture of normals as follows:

$$\varepsilon_i | v_i, \sigma \sim N((1-2p)v_i, 2\sigma v_i),\ v_i|\sigma \sim \text{Exp}(\sigma^{-1}p(1-p)),$$

where Exp(θ) denotes an exponential distribution with rate parameter θ. Denote $\mathbf{Y}$ as an $n \times 1$ response vector of y_i, $\mathbf{X}$ an $n \times k$ predictor matrix with ith row $\mathbf{x_i}^T$, we have

$$\mathbf{Y}|\boldsymbol{\beta}, \sigma, \mathbf{v}, \mathbf{X}, \boldsymbol{\Sigma} \sim N_n(\mathbf{X}\boldsymbol{\beta} + (1-2p)\mathbf{v}, 2\sigma\boldsymbol{\Sigma}),$$

where $\mathbf{v} = (v_1, \ldots, v_n)^T$ and $\boldsymbol{\Sigma}$ is the diagonal matrix of v_i. Given $\boldsymbol{\Sigma}$ and further let $\mathbf{Y}_p^* = \frac{1}{\sqrt{2}}(\mathbf{Y} - (1-2p)\mathbf{v})$, $\mathbf{X}^* = \frac{1}{\sqrt{2}}\mathbf{X}$ respectively, then $\mathbf{Y}_p^*$ follows a normal-type of conditional likelihood as

$$f(\mathbf{Y}_p^*|\boldsymbol{\beta}, \sigma, \mathbf{v}, \mathbf{X}^*, \boldsymbol{\Sigma}) \propto \sigma^{-n/2} \exp\{-\frac{1}{2\sigma}[\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta}]^T \boldsymbol{\Sigma}^{-1} [\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta}]\}. \quad (1)$$

3 *NIG* Prior and Posterior Distributions for Bayesian Quantile Regression

Mathematically, we introduce the definition of NIG [7] as follows.

Definition 1 Let $\boldsymbol{\beta}$ be a k-dimensional vector satisfying $-\infty < \boldsymbol{\beta} < \infty$ and $\delta > 0$ be the scalar parameter. The joint distribution of $(\boldsymbol{\beta}, \delta)$ follows the k-dimensional distribution $NIG_k(\boldsymbol{\mu}, \boldsymbol{\Lambda}, a, b)$ if

$$f(\boldsymbol{\beta}, \delta) = C\delta^{-(a+\frac{k}{2}+1)} \exp\{-\frac{1}{\delta}[b + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}(\boldsymbol{\beta} - \boldsymbol{\mu})]\},$$

where C is a proportionality constant. That is, $f(\delta)$ follows the inverse-gamma (IG) distribution with shape parameter a and scale parameter b, and $f(\boldsymbol{\beta}|\delta)$ follows the multivariate normal distribution with $k \times 1$ mean vector $\boldsymbol{\mu}$ and $k \times k$ precision matrix $\delta^{-1}\boldsymbol{\Lambda}$.

3.1 NIG Expression for Prior Distribution

Recall the likelihood function (1) of quantile regression and denote $\hat{\boldsymbol{\beta}}_p = (\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{Y}_p^*$, we can rewrite likelihood (1) as

$$
\begin{aligned}
f(\mathbf{Y}_p^*|\boldsymbol{\beta}, \sigma, \mathbf{v}, \mathbf{X}^*) \propto \sigma^{-\frac{n-k}{2}} & \exp\{-\frac{1}{2\sigma}[\mathbf{Y}_p^* - \mathbf{X}^*\hat{\boldsymbol{\beta}}_p]^T \boldsymbol{\Sigma}^{-1}[\mathbf{Y}_p^* - \mathbf{X}^*\hat{\boldsymbol{\beta}}_p]\} \\
& \sigma^{-\frac{k}{2}} \exp\{-\frac{1}{2\sigma}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_p)^T (\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_p)\} \\
& = (\sigma)^{-(a+\frac{k}{2}+1)} \exp\{-\frac{1}{\sigma}[b_p + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_p)^T \boldsymbol{\Lambda}(\boldsymbol{\beta} - \boldsymbol{\mu}_p)]\} \\
& \propto IG(a, b_p) N_k(\boldsymbol{\mu}_p, \sigma\boldsymbol{\Lambda}^{-1}), \qquad (2)
\end{aligned}
$$

where $\boldsymbol{\mu}_p = \hat{\boldsymbol{\beta}}_p$, $\boldsymbol{\Lambda} = \mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*$, $a = \frac{n-k-2}{2}$ and $b_p = \frac{1}{2}[\mathbf{Y}_p^* - \mathbf{X}^*\hat{\boldsymbol{\beta}}_p]^T\boldsymbol{\Sigma}^{-1}[\mathbf{Y}_p^* - \mathbf{X}^*\hat{\boldsymbol{\beta}}_p]$. According to Definition 1 with $\delta = \sigma$, the rewritten likelihood (2) can be represented as the structure of a k-dimensional distribution $NIG_k(\boldsymbol{\mu}, \boldsymbol{\Lambda}, a, b)$ in terms of parameters $(\boldsymbol{\beta}, \sigma)$.

Under the informative prior setting, following Alhamzawi and Yu [1], a conjugate prior for $(\boldsymbol{\beta}, \sigma)$ with a modification of Zellner's informative g-prior [28] in QR could be provided as

$$
\boldsymbol{\beta}|\sigma, \mathbf{v}, \mathbf{X}^*, \boldsymbol{\Sigma} \sim N_k(\mathbf{0}_k, g\sigma(\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*)^{-1}),\ f(\sigma) \propto \sigma^{-1},
$$

where $g > 0$ is a known scaling factor prescribed by the user. Smith and Kohn [20] proposed a Bayesian variable selection algorithm utilizing regression splines. They found that the choice of $g = 100$ works well and suggested to choose g between 10 and 1000. Following Smith and Kohn [20], the fixed setting of $g = 100$ has been considered by some other authors (see [8, 13], among others). Then we obtain the joint prior distribution of $(\boldsymbol{\beta}, \sigma)$

$$
f(\boldsymbol{\beta}, \sigma|\mathbf{v}, \mathbf{X}^*, \boldsymbol{\Sigma}) \propto \sigma^{-(\frac{k}{2}+1)} \exp\{-\frac{1}{\sigma}[\frac{1}{2}\boldsymbol{\beta}^T \frac{\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*}{g}\boldsymbol{\beta}]\}, \qquad (3)
$$

which is a special case of $NIG_k(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_{g0}, a_0, b_0)$ with $\boldsymbol{\mu}_0 = \mathbf{0}_k$, $\boldsymbol{\Lambda}_{g0} = \frac{\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*}{g}$, $a_0 = 0$, $b_0 = 0$.

3.2 *NIG Expression for Posterior Distribution*

The joint conditional posterior distribution $f(\boldsymbol{\beta}, \sigma, \mathbf{v}|\mathbf{Y}_p^*, \mathbf{X}^*)$ under the informative g-prior (3) is given by

$$
\begin{aligned}
f(\boldsymbol{\beta}, \sigma, \mathbf{v}|\mathbf{Y}_p^*, \mathbf{X}^*) &\propto f(\mathbf{Y}_p^*|\boldsymbol{\beta}, \sigma, \mathbf{v}) f(\boldsymbol{\beta}|\sigma, \mathbf{v}) f(\mathbf{v}|\sigma) f(\sigma) \\
&\propto \sigma^{-(\frac{3n+k+2}{2})} (\prod_{i=1}^n v_i^{-1/2}) |\mathbf{X}^{*T} \boldsymbol{\Sigma}^{-1} \mathbf{X}^*|^{1/2} \\
&\times \exp\{-\frac{1}{2\sigma}[(\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta}) \\
&+ \boldsymbol{\beta}^T \frac{\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*}{g} \boldsymbol{\beta} + 2p(1-p)\sum_{i=1}^n v_i]\}.
\end{aligned}
$$

Then the corresponding posterior $f(\boldsymbol{\beta}, \sigma|\mathbf{v}, \mathbf{Y}_p^*, \mathbf{X}^*)$ is given as follows:

$$
\begin{aligned}
f(\boldsymbol{\beta}, \sigma|\mathbf{v}, \mathbf{Y}_p^*, \mathbf{X}^*) &\propto \sigma^{-(\frac{3n+k+2}{2})} \exp\{-\frac{1}{2\sigma}[(\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta}) \\
&+ \boldsymbol{\beta}^T \frac{\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*}{g} \boldsymbol{\beta} + 2p(1-p)\sum_{i=1}^n v_i]\} \\
&= \sigma^{-(\frac{3n}{2}+\frac{k}{2}+1)} \exp\{-\frac{1}{\sigma}[\bar{b}_p + \frac{1}{2}(\boldsymbol{\beta} - \bar{\boldsymbol{\mu}}_p)^T \bar{\boldsymbol{\Lambda}} (\boldsymbol{\beta} - \bar{\boldsymbol{\mu}}_p)]\},
\end{aligned}
$$

which has an expression of $NIG_k(\bar{\boldsymbol{\mu}}_p, \bar{\boldsymbol{\Lambda}}, \bar{a}, \bar{b}_p)$, where $\bar{\boldsymbol{\mu}}_p = [(1+\frac{1}{g})\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*]^{-1}\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{Y}_p^*$, $\bar{\boldsymbol{\Lambda}} = (1+\frac{1}{g})\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*$, $\bar{a} = \frac{3n}{2}$, $\bar{b}_p = \frac{1}{2}\mathbf{Y}_p^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{Y}_p^* - \frac{1}{2}\bar{\boldsymbol{\mu}}_p^T \bar{\boldsymbol{\Lambda}}\bar{\boldsymbol{\mu}}_p + p(1-p)\sum_{i=1}^n v_i$. Moreover, the full conditional distributions of $\boldsymbol{\beta}$ and σ can be obtained respectively by

$$
f(\boldsymbol{\beta}|\sigma, \mathbf{v}, \mathbf{Y}_p^*, \mathbf{X}^*) \propto \exp\{-\frac{1}{2\sigma}[(\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta}) + \boldsymbol{\beta}^T \frac{\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*}{g} \boldsymbol{\beta}]\},
$$

which can be expressed as a k-dimensional normal $N_k(\bar{\boldsymbol{\mu}}_p, \sigma\bar{\boldsymbol{\Lambda}}^{-1})$, and

$$
\begin{aligned}
f(\sigma|\boldsymbol{\beta}, \mathbf{v}, \mathbf{Y}_p^*, \mathbf{X}^*) &\propto \sigma^{-(\frac{3n+k}{2}+1)} \exp\{-\frac{1}{2\sigma}[(\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta}) \\
&+ \boldsymbol{\beta}^T \frac{\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*}{g} \boldsymbol{\beta} + 2p(1-p)\sum_{i=1}^n v_i]\},
\end{aligned}
$$

which is an IG distribution with shape $\frac{3n+k}{2}$ and scale $\frac{1}{2}[(\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_p^* - \mathbf{X}^*\boldsymbol{\beta}) + \boldsymbol{\beta}^T \frac{\mathbf{X}^{*T}\boldsymbol{\Sigma}^{-1}\mathbf{X}^*}{g} \boldsymbol{\beta} + 2p(1-p)\sum_{i=1}^n v_i]$. The full posterior distribution of each $v_i, i = 1, 2, \ldots, n$ is also tractable:

$$
\begin{aligned}
f(v_i|\boldsymbol{\beta}, \sigma, y_i, \mathbf{x}_i) &\propto v_i^{-1} \exp\{-\frac{1}{4\sigma}[vi^{-1}((y_i - (1-2p)v_i - \mathbf{x}_i^T\boldsymbol{\beta})^2 + \frac{\boldsymbol{\beta}^T\mathbf{x}_i\mathbf{x}_i^T\boldsymbol{\beta}}{g})] - \frac{p(1-p)}{\sigma}v_i\} \\
&= v_i^{-1} \exp\{-\frac{1}{4\sigma}[v_i^{-1}((y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2 + \frac{\boldsymbol{\beta}^T\mathbf{x}_i\mathbf{x}_i^T\boldsymbol{\beta}}{g}) + v_i]\} \\
&= v_i^{-1} \exp\{-\frac{1}{2}(v_i^{-1}\bar{\xi}_i^2 + v_i\bar{\zeta}_i^2)\},
\end{aligned}
$$

where $\bar{\xi_i}^2 = [(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \boldsymbol{\beta}^T \mathbf{x}_i \mathbf{x}_i^T \boldsymbol{\beta}/g]/2\sigma$ and $\bar{\zeta_i}^2 = 1/2\sigma$, which can be recognized as a generalized inverse Gaussian distribution $GIG(0, \bar{\xi}_i, \bar{\zeta}_i)$ [2].

4 Big Data Based Algorithms for Bayesian Quantile Regression

4.1 *NIG* Multiplication Operator for Posterior Distribution

To derive the posterior distribution induced by the entire data set for Bayesian quantile regression, we first introduce the NIG multiplication operator defined as follows.

Proposition 1 *A general k-dimensional normal-inverse-gamma distribution $NIG_k(\boldsymbol{\mu}, \boldsymbol{\Lambda}, a, b)$ can be reformulated as a multiplication of H independent k-dimensional distributions $NIG_k(\boldsymbol{\mu}_h, \boldsymbol{\Lambda}_h, a_h, b_h), h = 1, \ldots, H$*

$$NIG_k(\boldsymbol{\mu}, \boldsymbol{\Lambda}, a, b) = \prod_{h=1}^{H} NIG_k(\boldsymbol{\mu}_h, \boldsymbol{\Lambda}_h, a_h, b_h), \tag{4}$$

where $\boldsymbol{\mu} = (\sum_{h=1}^H \boldsymbol{\Lambda}_h)^{-1} \sum_{h=1}^H \boldsymbol{\Lambda}_h \boldsymbol{\mu}_h$, $\boldsymbol{\Lambda} = \sum_{h=1}^H \boldsymbol{\Lambda}_h$, $a = \sum_{h=1}^H a_h + \frac{(H-1)(k+2)}{2}$ *and* $b = \sum_{h=1}^H b_h + \frac{1}{2} \sum_{h=1}^H (\boldsymbol{\mu}_h - \boldsymbol{\mu})^T \boldsymbol{\Lambda}_h (\boldsymbol{\mu}_h - \boldsymbol{\mu})$.

Recall the rewritten likelihood function of quantile regression (2) given in Sect. 3.1. If we partition the big data of $\mathbf{X}^*$ and $\mathbf{Y}_p^*$ into M subsets, where each $\mathbf{X}_m^*$ is an $n_m \times k$ matrix, $\mathbf{Y}_{pm}^*$ is an $n_m \times 1$ vector, $\boldsymbol{\Sigma}_m$ is an $n_m \times n_m$ diagonal block of $\boldsymbol{\Sigma}$ and $\sum_{m=1}^M n_m = n$, then the likelihood (2) can be reformulated as

$$f(\mathbf{Y}_p^* | \boldsymbol{\beta}, \sigma, \mathbf{v}, \mathbf{X}^*) \propto \sigma^{-\frac{\sum_{m=1}^M n_m - k}{2}} \exp\{-\frac{1}{2\sigma} \sum_{m=1}^M [\mathbf{Y}_{pm}^* - \mathbf{X}_m^* \hat{\boldsymbol{\beta}}_p]^T \boldsymbol{\Sigma}_m^{-1} [\mathbf{Y}_{pm}^* - \mathbf{X}_m^* \hat{\boldsymbol{\beta}}_p]\}$$
$$\sigma^{-\frac{k}{2}} \exp\{-\frac{1}{2\sigma} \sum_{m=1}^M (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_p)^T (\mathbf{X}_m^{*T} \boldsymbol{\Sigma}_m^{-1} \mathbf{X}_m^*)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_p)\},$$

which indicates a multiplication of M NIG distributions regarding parameters $(\boldsymbol{\beta}, \sigma)$

$$f(\mathbf{Y}_p^* | \boldsymbol{\beta}, \sigma, \mathbf{v}, \mathbf{X}^*) \propto \prod_{m=1}^M \sigma^{-(a_m^{(l)} + \frac{k}{2} + 1)} \exp\{-\frac{1}{\sigma}[b_m^{(l)} + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_p^{(l)})^T \boldsymbol{\Lambda}_m (\boldsymbol{\beta} - \boldsymbol{\mu}_p^{(l)})]\}$$
$$= \prod_{m=1}^M NIG(\boldsymbol{\mu}_p^{(l)}, \boldsymbol{\Lambda}_m^{(l)}, a_m^{(l)}, b_{pm}^{(l)}),$$

where the superscript (l) indicates the NIG parameters concerning $(\boldsymbol{\beta}, \sigma)$ for the likelihood function. $\boldsymbol{\mu}_p^{(l)} = \hat{\boldsymbol{\beta}}_p = (\sum_{m=1}^M \mathbf{X}_m^{*T} \boldsymbol{\Sigma}_m^{-1} \mathbf{X}_m^*)^{-1} \sum_{m=1}^M \mathbf{X}_m^{*T} \boldsymbol{\Sigma}_m^{-1} \mathbf{Y}_{pm}^*$, $\boldsymbol{\Lambda}_m^{(l)} =$

$\mathbf{X}_m^{*T}\boldsymbol{\Sigma}_m^{-1}\mathbf{X}_m^*$, $a_m^{(l)} = \frac{n_m-k-2}{2}$ and $b_{pm}^{(l)} = \frac{1}{2}[\mathbf{Y}_{pm}^* - \mathbf{X}_m^*\boldsymbol{\mu}_p^{(l)}]^T\boldsymbol{\Sigma}_m^{-1}[\mathbf{Y}_{pm}^* - \mathbf{X}_m^*\boldsymbol{\mu}_p^{(l)}]$. Then the full data posterior distribution is calibrated by the product of specified NIG prior and this multiplicative likelihood function, employing Eq. (4) with $H = M + 1$ in this case. The following Theorem 1 elaborates the acquisition of posterior distribution through the use of NIG multiplication operators.

Theorem 1 *Consider a linear quantile regression model with full big data observations* $\mathbf{X}$ *and* $\mathbf{Y}$. *Denote the posterior distribution of regression parameters* $(\boldsymbol{\beta}, \sigma)$, *under the prior* $NIG_k(\boldsymbol{\mu}^{(0)}, \boldsymbol{\Lambda}^{(0)}, a^{(0)}, b^{(0)})$, *be* $NIG_k(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Lambda}}, \bar{a}, \bar{b})$. *If we partition the whole data of size n into M subsets, each with an* $n_m \times k$ *matrix* $\mathbf{X}_m$ *and an* $n_m \times 1$ *vector* $\mathbf{Y}_m$, $m = 1, \ldots, M$, *and let* $\mathbf{X}_m^* = \frac{1}{\sqrt{2}}\mathbf{X}_m$, $\mathbf{Y}_{pm}^* = \frac{1}{\sqrt{2}}(\mathbf{Y}_m - (1-2p)\mathbf{v}_m)$, $\boldsymbol{\Sigma}_m = \mathrm{diag}(\mathbf{v_m})$, *where the latent variable* $\mathbf{v}_m$ *is an* $n_m \times 1$ *vector generated from the exponential distribution with rate* $\sigma^{-1}p(1-p)$, *then the full data posterior distribution can be formulated as*

$$NIG_k(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Lambda}}, \bar{a}, \bar{b}) = NIG_k(\boldsymbol{\mu}^{(0)}, \boldsymbol{\Lambda}^{(0)}, a^{(0)}, b^{(0)}) \prod_{m=1}^{M} NIG_k(\boldsymbol{\mu}_p^{(l)}, \boldsymbol{\Lambda}_m^{(l)}, a_m^{(l)}, b_{pm}^{(l)}),$$

where $\bar{\boldsymbol{\mu}} = (\boldsymbol{\Lambda}^{(0)} + \sum_{m=1}^M \mathbf{X}_m^{*T}\boldsymbol{\Sigma}_m^{-1}\mathbf{X}_m^*)^{-1}(\boldsymbol{\Lambda}^{(0)}\boldsymbol{\mu}^{(0)} + \sum_{m=1}^M \mathbf{X}_m^{*T}\boldsymbol{\Sigma}_m^{-1}\mathbf{Y}_{pm}^*)$, $\bar{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}^{(0)} + \sum_{m=1}^M \mathbf{X}_m^{*T}\boldsymbol{\Sigma}_m^{-1}\mathbf{X}_m^*$, $\bar{a} = a^{(0)} + \frac{n}{2}$ *and* $\bar{b} = b^{(0)} + \frac{1}{2}[\sum_{m=1}^M \mathbf{Y}_{pm}^{*T}\boldsymbol{\Sigma}_m^{-1}\mathbf{Y}_{pm}^* + \boldsymbol{\mu}^{(0)T}\boldsymbol{\Lambda}^{(0)}\boldsymbol{\mu}^{(0)} - \bar{\boldsymbol{\mu}}^T\bar{\boldsymbol{\Lambda}}\bar{\boldsymbol{\mu}}]$.

4.2 Algorithms for Bayesian Quantile Regression

Consider the linear QR model for the p-th quantile $(0 < p < 1)$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\mathbf{Y}$ is an $n \times 1$ response vector, $\mathbf{X}$ is an $n \times k$ predictor matrix, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of $ALD(0, \sigma, p)$ disturbances. Then model (5) is equivalent to

$$\mathbf{Y}_p^* = \mathbf{X}^*\boldsymbol{\beta} + \sqrt{\sigma}\boldsymbol{\varepsilon}^*,$$

where $\mathbf{Y}_p^* = \frac{1}{\sqrt{2}}(\mathbf{Y} - (1-2p)\mathbf{v})$, $\mathbf{X}^* = \frac{1}{\sqrt{2}}\mathbf{X}$ and $\boldsymbol{\varepsilon}^* \sim N_n(\mathbf{0}_n, \boldsymbol{\Sigma})$ with $n \times n$ known positive definite covariance matrix $\boldsymbol{\Sigma}$. Then we proceed to Bayesian inference for big data quantile regressions through the proposed NIG multiplication operator. We consider model (5) under the g-prior (3) for $(\boldsymbol{\beta}, \sigma)$, and partition the entire data set into M subsets $(\mathbf{X}_m, \mathbf{Y}_m)$ with individual sample size n_m, $m = 1, \ldots, M$. Then the posterior distribution for the whole data can be obtained by merging the given prior with the multiplication of M subset NIG distributions induced from the massive observations. Based on this, an efficient divide-and-conquer algorithm for big data Bayesian quantile regression is provided as below.

Algorithm 1 Consider a pth $(0 < p < 1)$ Bayesian quantile regression under g-prior (3) with the observed $n \times k$ design matrix $\mathbf{X}$ and $n \times 1$ response vector $\mathbf{Y}$, where the large data set cannot be fit into a single computer due to the memory constraint. We can obtain the full data posterior distribution by the following divide-and-conquer algorithm.

Step 1 partition the entire data set into M subsets $\mathbf{X}_m, \mathbf{Y}_m, m = 1, 2, \ldots, M$, where $\mathbf{X}_m$ is an $n_m \times k$ matrix, $\mathbf{Y}_m$ is an $n_m \times 1$ vector and $\sum_{m=1}^{M} n_m = n$.

Step 2 for each subset $\mathbf{X}_m, \mathbf{Y}_m$, a Gibbs sampler for sampling $\boldsymbol{\beta}_m, \sigma_m$ and $\mathbf{v}_m$ in the BQR would follow the sub-steps presented below:

2.1 denote j as the iteration count. Then set $j = 0$ and establish $(\boldsymbol{\beta}_m^{(j=0)}, \sigma_m^{(j=0)}, \mathbf{v}_m^{(j=0)})$ to some starting values.

2.2 follow the full conditional distributions of $\boldsymbol{\beta}_m, \sigma_m$ and $\mathbf{v}_m$,

(i) sample $\mathbf{v}_m^{(j+1)}$ from $f(\mathbf{v}_m|\boldsymbol{\beta}_m^{(0)}, \sigma_m^{(0)})$.

(ii) sample $\sigma_m^{(j+1)}$ from $f(\sigma_m|\boldsymbol{\beta}_m^{(0)}, \mathbf{v}_m^{(1)})$.

(iii) sample $\boldsymbol{\beta}_m^{(j+1)}$ from $f(\boldsymbol{\beta}_m|\sigma_m^{(1)}, \mathbf{v}_m^{(1)})$.

2.3 set $j = j + 1$ and return to **Step 2.2** until $j = L$, where L is the number of iteration times.

Step 3 calculate the empirical estimates of the means $\bar{\boldsymbol{\beta}}_m$ and $\bar{\sigma}_m$ separately based on the $(L - B)$ realizations of the Gibbs sequence (discarding the first B iterations as a burn-in). Then generate an n_m i.i.d. sample on v_i, where $v_i \sim GIG(0, \bar{\xi}_i, \bar{\zeta}_i)$, with $\bar{\xi}_i^{\,2} = [(y_i - \mathbf{x}_i^T \bar{\boldsymbol{\beta}}_m)^2 + \bar{\boldsymbol{\beta}}_m^T \mathbf{x}_i \mathbf{x}_i^T \bar{\boldsymbol{\beta}}_m / g]/2\bar{\sigma}_m$ and $\bar{\zeta}_i^{\,2} = 1/2\bar{\sigma}_m, i = 1, 2, \ldots, n_m$. Let $\mathbf{X}_m^* = \frac{1}{\sqrt{2}}\mathbf{X}_m$, $\mathbf{Y}_{pm}^* = \frac{1}{\sqrt{2}}(\mathbf{Y}_m - (1 - 2p)\mathbf{v}_m)$, where $\mathbf{v}_m$ is the corresponding $n_m \times 1$ vector of v_i for each subset, and denote $\boldsymbol{\Sigma}_m$ as an $n_m \times n_m$ diagonal matrix with $\mathbf{v}_m$ its diagonal vector, $m = 1, 2, \ldots, M$.

Step 4 for each subset, the corresponding likelihood can be represented as a form of $NIG_k(\boldsymbol{\mu}_{pm}, \boldsymbol{\Lambda}_m, a_m, b_{pm})$ distribution for $(\boldsymbol{\beta}, \sigma)$. Obtain the multiplicative distribution $NIG_k(\boldsymbol{\mu}_p, \boldsymbol{\Lambda}, a, b_p) = \prod_{m=1}^{M} NIG(\boldsymbol{\mu}_{pm}, \boldsymbol{\Lambda}_m, a_m, b_{pm})$, then the full data posterior can be given by merging the g-prior $NIG_k(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_{g0}, a_0, b_0)$ and distribution $NIG_k(\boldsymbol{\mu}_p, \boldsymbol{\Lambda}, a, b_p)$:

$$NIG_k(\bar{\boldsymbol{\mu}}_p, \bar{\boldsymbol{\Lambda}}, \bar{a}, \bar{b}_p) = NIG_k(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_{g0}, a_0, b_0) NIG_k(\boldsymbol{\mu}_p, \boldsymbol{\Lambda}, a, b_p),$$

where $\bar{\boldsymbol{\mu}}_p = [(1 + \frac{1}{g}) \sum_{m=1}^{M} \mathbf{X}_m^{*T} \boldsymbol{\Sigma}_m^{-1} \mathbf{X}_m^*)]^{-1} \sum_{m=1}^{M} \mathbf{X}_m^{*T} \boldsymbol{\Sigma}_m^{-1} \mathbf{Y}_{pm}^*$, $\bar{\boldsymbol{\Lambda}} = (1 + \frac{1}{g}) \sum_{m=1}^{M} \mathbf{X}_m^{*T} \boldsymbol{\Sigma}_m^{-1} \mathbf{X}_m^*$, $\bar{a} = \frac{3n}{2}$, $\bar{b}_p = \frac{1}{2}[\sum_{m=1}^{M} \mathbf{Y}_{pm}^{*T} \boldsymbol{\Sigma}_m^{-1} \mathbf{Y}_{pm}^* - \bar{\boldsymbol{\mu}}_p^T \bar{\boldsymbol{\Lambda}} \bar{\boldsymbol{\mu}}_p] + p(1 - p) \sum_{m=1}^{M} \|\mathbf{v_m}\|_1$ and $\|\cdot\|_1$ denotes the ℓ_1 norm of a vector.

Table 1 Summary statistics for wind power observations at Aeolos, Iweco and Rokas

	Aeolos	Iweco	Rokas
Min	0.000	0.000	0.000
Quantile (0.25)	1.692	0.921	1.573
Median	4.002	2.112	4.579
Mean	4.142	2.141	4.857
Quantile (0.75)	6.745	3.426	8.049
Max	8.302	4.549	11.635
Standard deviation	2.649	1.346	3.407
Sample size	17,819	15,621	21,949

5 Real-Data Analysis

In this section, we illustrate our divide-and-conquer algorithm for big data Bayesian quantile regression by a real-world data analysis. We use hourly wind power data recorded from 31 December 2007 to 30 December 2010 at the following three wind farms in Crete: Aeolos, Iweco and Rokas. The data is a collection of hourly observations for wind speed (measured in m/s), direction (measured in degrees) and power (measured in megawatts). A complete wind power data of the year 2010 is examined in Taylor [21]. We remove all the missing data and retain positive observations of the recorded hourly periods. Table 1 presents the summary statistics for wind power observations (in MW) at Aeolos, Iweco and Rokas respectively.

We fit our big data BQR by modeling the wind power as a linear function of wind speed and direction. We implement Algorithm 1 for these three power sequences at $p = 0.50$ and $p = 0.95$ respectively. In each case, the Gibbs samplers are run for 11000 iterations, discarding the first 1000 as a burn-in. For Aeolos farm, the whole observations are partitioned into 50 subsets with the size of $n_1 = n_2 \ldots = n_{49} = 356$ and $n_{50} = 375$. For Iweco, we partition the whole data into 50 subsets with the size of $n_1 = n_2 \ldots = n_{49} = 312$ and $n_{50} = 333$. For Rokas, we consider 50 subsets as $n_1 = n_2 \ldots = n_{49} = 438$ and $n_{50} = 487$. We assign the informative g-prior by choosing $g = 100$. Table 2 displays the estimates and posterior standard deviations in our big data BQR model for the given three wind power series separately. Note that for all power series, the estimated coefficients of direction are close to zero at the measured percentiles, meaning that the effect of wind direction on power seems to be minor. Instead, wind power presents a much stronger correlation to speed than to direction. The positive coefficients of speed indicate that as wind speed increases, so does the power capacity. Furthermore, it is visible that speed has a greater impact on higher (95th percentile) power than lower (50th percentile) power capacity for all the three aforementioned wind farms.

Table 2 Coefficient estimates along with posterior standard deviations (S.D.) for Aeolos, Iweco and Rokas in big data BQR analysis

Model covariates	Aeolos				Iweco				Rokas			
	$p = 0.50$		$p = 0.95$		$p = 0.50$		$p = 0.95$		$p = 0.50$		$p = 0.95$	
	Coeff.	S.D.	Coeff.	S.D.	Coeff.	S.D.	Coeff.	S.D.	Coeff.	S.D.	Coeff.	S.D.
Intercept	−2.8624	0.0151	−3.6681	0.0160	−0.7907	0.0110	−0.5663	0.0135	−2.8004	0.0130	−1.9270	0.0150
Speed	0.7485	0.0151	1.0447	0.0018	0.3770	0.0015	0.5316	0.0015	0.7860	0.0013	1.0616	0.0010
Direction	−0.0003	0.0000	−0.0021	0.0000	−0.0023	0.0000	−0.0039	0.0000	0.0005	0.0000	−0.0040	0.0000

6 Summary and Conclusion

This paper extends the divide-and-conquer algorithm for big data analysis from traditional mean-based linear regression to quantile regression under Bayesian perspectives. This is achieved by using ALD-based working likelihood functions and conjugate NIG priors. The resulting algorithms are easily implemented and the real-data illustrations present that wind speed has a greater impact on higher power values than lower ones, showing the proposed methods are promising. The developed algorithms can be investigated for other energy-related observations within big data scenario, such as solar radiation and electrical power demand series. In this empirical study, we have assigned the positive scaling g-prior by fixing it to be the experimental value $g = 100$, as suggested in Smith and Kohn [20] after extensive testing. However, a potential alternative is to assign a hyper-prior distribution on the g parameter rather than keep it as a fixed constant. Under such circumstances, the unknown parameter g can be estimated from the available data. Moreover, the undesirable "Information Paradox", which relates to the limiting behavior of the Bayes factor for model selection with fixed g, can be avoided (see [14, 16]). Our possible future work will focus on developing a novel Bayesian quantile regression for fitting single-index models under high-dimensional data context, and its penalized version for efficient variable selection implementations.

Acknowledgements The authors would like to thank the support of the National Social Science Foundation of China (Series number: 21BTJ040).

References

1. Alhamzawi, R., Yu, K.: Conjugate priors and variable selection for Bayesian quantile regression. Comput. Stat. Data Anal. **64**, 209–219 (2013)
2. Barndorff-Nielsen, O.E., Shephard, N.: Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. J. Roy. Stat. Soc. B: Stat. Methodol. **63**, 167–241 (2001)
3. Briollais, L., Durrieu, G.: Application of quantile regression to recent genetic and -omic studies. Hum. Genet. **133**, 951–966 (2014)
4. Bernardi, M., Gayraud, G., Petrella, L.: Bayesian tail risk interdependence using quantile regression. Bayesian Anal. **10**, 553–603 (2015)
5. Cole, T.J., Green, P.J.: Smoothing reference centile curves: the LMS method and penalized likelihood. Stat. Med. **11**, 1305–1319 (1992)
6. Chen, X., Liu, W., Zhang, Y.: Quantile regression under memory constraint. Ann. Stat. **47**, 3244–3273 (2019)
7. Denison, D.G., Holmes, C.C., Mallick, B.K., Smith, A.F.: Bayesian Methods for Nonlinear Classification and Regression. Wiley, Hoboken (2002)
8. Gupta, M., Qu, P., Ibrahim, J.G.: A temporal hidden Markov regression model for the analysis of gene regulatory networks. Biostatistics **8**, 805–820 (2007)
9. Gu, Y., Fan, J., Kong, L., Ma, S., Zou, H.: ADMM for high-dimensional sparse penalized quantile regression. Technometrics **60**, 319–331 (2018)
10. Gonçalves, K.C., Migon, H.S., Bastos, L.S.: Dynamic quantile linear models: a Bayesian approach. Bayesian Anal. **15**, 335–362 (2020)

11. Koenker, R., Hallock, K.F.: Quantile regression: an introduction. J. Econ. Perspect. **15**, 143–156 (2001)
12. Kozumi, H., Kobayashi, G.: Gibbs sampling methods for Bayesian quantile regression. J. Stat. Comput. Simul. **81**, 1565–1578 (2011)
13. Lee, K.E., Sha, N., Dougherty, E.R., Vannucci, M., Mallick, B.K.: Gene selection: a Bayesian variable selection approach. Bioinformatics **19**, 90–97 (2003)
14. Liang, F., Paulo, R., Molina, G., Clyde, M.A., Berger, J.O.: Mixtures of g priors for Bayesian variable selection. J. Am. Stat. Assoc. **103**, 410–423 (2008)
15. Lum, K., Gelfand, A.E.: Spatial quantile multiple regression using the asymmetric Laplace process. Bayesian Anal. **7**, 235–258 (2012)
16. Perrakis, K., Ntzoufras, I.: Bayesian variable selection using the hyper-g prior in WinBUGS. Wiley Interdisc. Rev. Comput. Stat. **10**, e1442 (2018)
17. Petrella, L., Raponi, V.: Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. J. Multivar. Anal. **173**, 70–84 (2019)
18. Reed, C., Yu, K.: A partially collapsed Gibbs sampler for Bayesian quantile regression (2009)
19. Rodrigues, T., Fan, Y.: Regression adjustment for noncrossing Bayesian quantile regression. J. Comput. Graph. Stat. **26**, 275–284 (2017)
20. Smith, M., Kohn, R.: Nonparametric regression using Bayesian variable selection. J. Econom. **75**, 317–343 (1996)
21. Taylor, J.W.: Probabilistic forecasting of wind power ramp events using autoregressive logit models. Eur. J. Oper. Res. **259**, 703–712 (2017)
22. Wu, Y., Yin, G.: Conditional quantile screening in ultrahigh-dimensional heterogeneous data. Biometrika **102**, 65–76 (2015)
23. Wang, Y., Feng, X.N., Song, X.Y.: Bayesian quantile structural equation models. Struct. Equ. Model. **23**, 246–258 (2016)
24. Yu, K., Moyeed, R.A.: Bayesian quantile regression. Stat. Probab. Lett. **54**, 437–447 (2001)
25. Yu, K., Lu, Z., Stander, J.: Quantile regression: applications and current research areas. J. Roy. Stat. Soc. Ser. D Stat. **52**, 331–350 (2003)
26. Yu, K., Stander, J.: Bayesian analysis of a Tobit quantile regression model. J. Econ. **137**, 260–276 (2007)
27. Yu, L., Lin, N., Wang, L.: A parallel algorithm for large-scale nonconvex penalized quantile regression. J. Comput. Graph. Stat. **26**, 935–939 (2017)
28. Zellner, A.: On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: Goel P.K., Zellner, A. (eds.) Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, pp. 233–243. Elsevier, North-Holland (1986)