



Bayesian composite L^p -quantile regression

Lukas Arnroth¹ 

Received: 23 January 2023 / Accepted: 19 January 2024
© The Author(s) 2024

Abstract

L^p -quantiles are a class of generalized quantiles defined as minimizers of an asymmetric power function. They include both quantiles, $p = 1$, and expectiles, $p = 2$, as special cases. This paper studies composite L^p -quantile regression, simultaneously extending single L^p -quantile regression and composite quantile regression. A Bayesian approach is considered, where a novel parameterization of the skewed exponential power distribution is utilized. Further, a Laplace prior on the regression coefficients allows for variable selection. Through a Monte Carlo study and applications to empirical data, the proposed method is shown to outperform Bayesian composite quantile regression in most aspects.

Keywords Skewed exponential power distribution · L^p -quantile regression · Markov chain Monte Carlo

1 Introduction

Conventional regression analysis, based on the ordinary least squares (OLS) framework, plays a vital role in exploring the relationship amongst variables. It is, however, well known that the OLS estimator is not robust to deviations from normality of the response in the form of heavy-tailed distributions or outliers. As a robust alternative, quantile regression (QR), introduced by Koenker and Bassett (1978), has become a popular paradigm to describe a more complete conditional distribution information. One caveat is that QR can have arbitrarily small relative efficiency compared to the OLS estimator based on a single quantile. Further, the quantile regression at one quantile may provide more efficient estimates than for another quantile. Zou and Yuan (2008) proposed the simultaneous estimation over multiple quantiles, with equal weights assigned to each quantile considered, to abate the issues of the single quantile regression. Zhao and Xiao (2014) extended the composite quantile regression to

✉ Lukas Arnroth
lukas.arnroth@statistik.uu.se

¹ Department of Statistics, Uppsala University, Uppsala, Sweden

unequal weights for each quantile. A Bayesian estimation procedure for the composite model with unequal weights was outlined by Huang and Chen (2015).

A different set of issues with QR regression is from the perspective of axiomatic theory. Artzner et al. (1999) provide a foundation of coherent risk measures, where it's shown that quantiles are not coherent as they don't satisfy the criterion of subadditivity, see, e.g. Bellini et al. (2014) for more details. Generalized quantiles represent a separate extension of QR, which addresses the issues with quantiles as a risk measure, and is based on modifying the asymmetric linear loss function of quantiles by considering more general loss functions. Newey and Powell (1987) introduced expectiles as minimization of an asymmetric squared function. Chen (1996) introduced L^p -quantiles as the minimization of an asymmetric power function, which includes both the expectile and quantile loss functions as special cases. As L^p -quantiles are coherent risk measures, they have gained much recent popularity in actuarial applications and extreme value analysis (Bellini et al. 2014; Usseglio-Carleve 2018; Daouia et al. 2019; Konen and Paindaveine 2022). Apart from coherency, the usage of L^p -quantiles is mainly motivated by their flexibility, bridging between the robustness of quantiles and the sensitivity of expectiles. However, the L^p -quantile approach is not without drawbacks, mainly that L^p -quantiles does not have an interpretation as direct as ordinary quantiles. For a general discussion on relating the interpretation of L^p -quantiles to ordinary quantiles, see Jones (1994).

This paper aims to extend L^p -quantile regression to composite L^p -quantile regression from the Bayesian perspective, emulating the extension for ordinary quantiles by Huang and Chen (2015). Bayesian single L^p -quantile regression, based on the Skewed exponential power distribution (SEPD) (Komunjer 2007; Zhu and Zinde-Walsh 2009), has been considered by Bernardi et al. (2018); Arnroth and Vegelius (2023). However, maximization of the likelihood of the SEPD corresponds to the minimization of a transformation of the L^p -quantile loss function. Therefore, a novel parametrization of the SEPD, based directly on the loss function of L^p -quantiles, is introduced. Compared to the single L^p -quantile setting, the composite extension significantly alleviates the issues of interpretation as the parameters of interest for the Bayesian composite L^p -quantile regression (BCLQR) are the same as those of Bayesian composite quantile regression (BCQR). Hence, results based on L^1 -quantiles and L^p -quantiles will be directly comparable. Furthermore, following Huang and Chen (2015), the issue of variable selection is simultaneously considered by considering a Laplace prior for the regression coefficients, which translates into a L^1 penalty on the regression coefficients, i.e., Lasso (Tibshirani 1996).

The article is organized as follows. In Sect. 2, the BCLQR method with a lasso penalty on the regression coefficients is introduced. Section 3 presents numerical results with applications to both simulated and empirical data. Finally, discussion and conclusions are put in Sect. 4.

2 Bayesian composite L^p -quantile regression

Consider the following linear model

$$y = b_0 + \mathbf{x}^T \boldsymbol{\beta} + \varepsilon, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^m$ is the m -dimensional covariate, $\boldsymbol{\beta} \in \mathbb{R}^m$ is the m -dimensional vector of unknown parameters, $b_0 \in \mathbb{R}$ is the intercept, $y \in \mathbb{R}$ is the response, and ε is the noise. The conditional τ th L^p -quantile of $y|\mathbf{x}$ is

$$b_0 + \mathbf{x}^T \boldsymbol{\beta} + q_\tau = b_\tau + \mathbf{x}^T \boldsymbol{\beta},$$

where q_τ is the τ th L^p -quantile of ε for $\tau \in (0, 1)$, independent of \mathbf{x} . For a random variable z with cumulative density function F , the τ th L^p -quantile is defined as (Chen 1996)

$$q_\tau = \arg \min_q \int \rho_{\tau,p}(z - q) dF(z), \tag{2}$$

where $\rho_{\tau,p}(\cdot)$ is an asymmetric power function defined as

$$\rho_{\tau,p}(y) = |\tau - I(y \leq 0)| |y|^p. \tag{3}$$

For a random sample $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$, the τ th L^p -quantile regression model estimates b_τ and $\boldsymbol{\beta}$ by solving

$$(\hat{b}_\tau, \hat{\boldsymbol{\beta}}) = \operatorname{argmin}_{b_\tau, \boldsymbol{\beta}} \sum_{i=1}^n \rho_{\tau,p}(y_i - b_\tau - \mathbf{x}_i^T \boldsymbol{\beta}), \tag{4}$$

By setting $p = 1$ in (4), we retain the standard quantile regression estimator. To recast minimization of (4) into maximization of a likelihood function, we assume the noise term in (1) to follow a density of the form

$$f(y; \tau, p) = K_{\tau,p} \exp\{-|\tau - I(y \leq 0)| |y|^p\} = K_{\tau,p} \exp\{-\rho_{\tau,p}(y)\}, \tag{5}$$

where $p > 0$ and $K_{\tau,p}^{-1} = \Gamma(1 + 1/p)(\tau^{-1/p} + (1 - \tau)^{-1/p})$. The density function (5) corresponds to a re-scaling of the skewed exponential power distribution (Komunjer 2007; Zhu and Zinde-Walsh 2009). Furthermore, setting $p = 1$ directly retains the asymmetric Laplace distribution utilized for Bayesian quantile regression (Kozumi and Kobayashi 2011). For scale $\sigma > 0$ and location $\mu \in \mathbb{R}$, the density is

$$p(y; \mu, \sigma, \tau, p) = \sigma^{-1} f\left(\sigma^{-1}(y - \mu); \tau, p\right). \tag{6}$$

The reparametrization $\eta = \sigma^p$ shall be considered in the sequel, with a prior placed directly on η , which simplifies the subsequent MCMC sampling scheme. Following

Huang and Chen (2015) and Zhao and Xiao (2014), composite L^p -quantile regression is defined as

$$(\hat{b}_{\tau_1}, \dots, \hat{b}_{\tau_K}, \hat{\boldsymbol{\beta}}) = \operatorname{argmin}_{b_{\tau_1}, \dots, b_{\tau_K}, \boldsymbol{\beta}} \sum_{i=1}^n \left\{ \sum_{k=1}^K w_k \rho_{\tau_k, p}(y_i - b_{\tau_k} - \mathbf{x}_i^T \boldsymbol{\beta}) \right\}, \quad (7)$$

where $0 < \tau_1 < \dots < \tau_K < 1$ are quantile levels and $0 < w_k < 1$ is the weight of the k th component, where $\mathbf{w} = (w_1, \dots, w_K)^T$ is defined on the $(K - 1)$ -dimensional simplex $\{\mathbf{w} \in [0, 1]^K : \sum_{i=1}^K w_k = 1\}$. Note that, unlike fitting K independent quantile models, (7) assumes the same $\boldsymbol{\beta}$ across quantiles. By fixing $p = 1$ in (7), the estimation procedure of Huang and Chen (2015) is retained.

The joint distribution of $\mathbf{y} = (y_1, \dots, y_n)^T$ given $\mathbf{X} = (\mathbf{x}_1^T, \dots, \mathbf{x}_n^T)^T$ for a composite model is

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{w}, p, \sigma) = \prod_{i=1}^n \left(\sum_{k=1}^K w_k p_{\tau_k}(y_i|\mathbf{x}_i, b_{\tau_k}, \boldsymbol{\beta}, \eta, p) \right), \quad (8)$$

where $\mathbf{b} = (b_{\tau_1}, \dots, b_{\tau_K})^T$ and

$$p_{\tau_k}(y_i|\mathbf{x}_i, b_{\tau_k}, \boldsymbol{\beta}, \sigma, p) = \frac{K \tau_k, p}{\eta^{1/p}} \exp \left\{ -\frac{1}{\eta} \rho_{\tau_k, p}(y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k}) \right\}.$$

To solve (8), we introduce a matrix \mathbf{C} for cluster assignment whose i, k th element, C_{ik} , is equal to 1 if the i th subject belongs to the k th cluster, otherwise $C_{ik} = 0$. \mathbf{C} is treated as a missing value and we start from the complete likelihood which has the form

$$\begin{aligned} p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{w}, \sigma, p, \boldsymbol{\tau}, \mathbf{C}) &= \prod_{i=1}^n \prod_{k=1}^K [w_k p_{\tau_k}(y_i|\mathbf{x}_i, b_{\tau_k}, \boldsymbol{\beta}, \sigma, p)]^{C_{ik}} \\ &= \frac{1}{\eta^{n/p} \Gamma(1 + 1/p)^n} \prod_{k=1}^K \left(\frac{w_k}{\tau_k^{1/p} + (1 - \tau_k)^{1/p}} \right)^{n_k} \\ &\quad \exp \left\{ -\frac{1}{\eta} \sum_{i=1}^n \sum_{k=1}^K C_{ik} \rho_{\tau_k, p}(y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k}) \right\}, \end{aligned}$$

where $n_k = \sum_{i=1}^n C_{ik}$.

To perform variable selection, we consider the standard Laplace prior for $\boldsymbol{\beta}$

$$\pi(\boldsymbol{\beta}) = \frac{1}{2^m} \exp \left\{ -\sum_{j=1}^m |\beta_j| \right\}.$$

A Dirichlet prior is considered for \mathbf{w} ,

$$\pi(\mathbf{w}) = \text{Dirichlet}(\alpha_1, \dots, \alpha_K),$$

with $\alpha_1 = \dots = \alpha_K = 0.1$. The priors of σ and p are set to

$$\pi(\eta) = \mathcal{IG}(0.1, 0.1) \quad \text{and} \quad \pi(p) = \mathcal{U}(0, 5),$$

where $\mathcal{IG}(a, b)$ denotes the inverse Gamma distribution with shape a and scale b and $\mathcal{U}(a, b)$ denotes the continuous Uniform distribution on $[a, b]$. Here, the upperbound on p was set to 5 which was sufficiently high to not impact the posterior samples of p in Sect. 3 meaningfully. Also, from a practical perspective, limiting the values of p to not be too large is motivated by the increased flatness around the mode of the SEPD as p increases, which could cause issues for a MCMC procedure. It's worth noting that, from a practitioner's perspective, the most intuitive approach is to restrict the prior on p to a continuous distribution within the range $p \in [1, 2]$. As such, the interpretation of the procedure is along the lines of a bridge between the robustness of quantiles and the sensitivity of expectiles. We avoid such restrictions here however to allow for increased flexibility of the estimation procedure. Additionally, we treat the quantile specific intercepts as described in Huang and Chen (2015), setting $\pi(\mathbf{b}) \propto 1$. Such an improper prior yields a proper posterior for the single L^p -quantile regression (Arnroth and Vegelius 2023). The posterior distribution is thus given by

$$\pi(\boldsymbol{\beta}, \mathbf{b}, \eta, p | \mathbf{y}, \mathbf{X}) \propto p(\mathbf{y} | \mathbf{X}, \mathbf{b}, \eta, p, \boldsymbol{\tau}, \mathbf{C}) \pi(\boldsymbol{\beta}) \pi(\eta) \pi(\mathbf{w}) \pi(p). \tag{9}$$

A MCMC procedure is utilized to sample from (9). The conditional distribution of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \mathbf{b}, \mathbf{C}, \mathbf{w}, \eta, p, \lambda) \propto \exp \left\{ -\frac{1}{\eta} \sum_{i=1}^n \sum_{k=1}^K C_{ik} \rho_{\tau_k, p} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k} \right) - \lambda \sum_{j=1}^m |\beta_j| \right\}, \tag{10}$$

of which the normalizing constant is unknown. Due to the possible large dimension of $\boldsymbol{\beta}$, the Metropolis adjusted Langevin algorithm (MALA) is used to sample (10) efficiently. Proposals are generated as

$$\boldsymbol{\beta}' = \boldsymbol{\beta} + \frac{\epsilon_\beta^2}{2} \mathbf{A} \nabla_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}) + \epsilon_\beta \sqrt{\mathbf{A}} \mathbf{Z},$$

where $\mathcal{L}(\boldsymbol{\beta})$ is the log of (10), $\mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1}$, and $\mathbf{Z} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I})$. Note that $\mathbf{X}^T \mathbf{X}$ is used rather than the Hessian due to issues with positive definiteness for some values

of p . Ignoring non-differentiability of $|\cdot|$ at 0, the gradient is

$$\begin{aligned} \nabla_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}) = & \frac{p}{\eta} \sum_{k=1}^K \left[(1 - \tau_k) \sum_{i \in N_{1,k}} |\mathbf{x}_i^T \boldsymbol{\beta} + b_{\tau_k} - y_i|^{p-1} \mathbf{x}_i \right. \\ & \left. - \tau_k \sum_{i \in N_{2,k}} C_{ik} |y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k}|^{p-1} \mathbf{x}_i \right] - \lambda, \end{aligned}$$

where $N_{1,k} = \{i : y_i < \mathbf{x}_i^T \boldsymbol{\beta} + b_{\tau_k} \text{ and } C_{ik} = 1\}$ and $N_{2,k} = \{i : y_i > \mathbf{x}_i^T \boldsymbol{\beta} + b_{\tau_k} \text{ and } C_{ik} = 1\}$. The acceptance probability is then given by

$$\varphi_{\boldsymbol{\beta}}(\boldsymbol{\beta}', \boldsymbol{\beta}) = 1 \wedge \frac{\pi(\boldsymbol{\beta}' | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \mathbf{b}, \mathbf{C}, \mathbf{w}, \eta, p, \lambda) q(\boldsymbol{\beta} | \boldsymbol{\beta}')}{\pi(\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \mathbf{b}, \mathbf{C}, \mathbf{w}, \eta, p, \lambda) q(\boldsymbol{\beta}' | \boldsymbol{\beta})},$$

where $q(\boldsymbol{\beta} | \boldsymbol{\beta}') = \mathcal{N}(\boldsymbol{\beta} | \boldsymbol{\beta}' + \epsilon_{\boldsymbol{\beta}}^2 / 2 \mathbf{A} \nabla_{\boldsymbol{\beta}'} \mathcal{L}(\boldsymbol{\beta}'), \epsilon_{\boldsymbol{\beta}}^2 \mathbf{A})$ and $a \wedge b = \min\{a, b\}$.

The conditional distribution for the LP-quantile specific intercepts is

$$\begin{aligned} \pi(\mathbf{b} | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{C}, \mathbf{w}, \eta, p) & \propto \\ \exp \left\{ -\frac{1}{\eta} \sum_{i=1}^n \sum_{k=1}^K C_{ik} \rho_{\tau_k, p} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_k \right) \right\}. \end{aligned} \quad (11)$$

As for $\boldsymbol{\beta}$, the MALA is considered using only the first order derivative of (11). Proposals are thus generated as

$$\mathbf{b}' = \mathbf{b} + \frac{\epsilon_{\mathbf{b}}^2}{2} \nabla_{\mathbf{b}} \mathcal{L}(\mathbf{b}) + \epsilon \mathbf{Z},$$

where $\nabla_{\mathbf{b}} \mathcal{L}(\mathbf{b})$ has k th element

$$\begin{aligned} \frac{\partial}{\partial b_{\tau_k}} \mathcal{L}(\mathbf{b}) = & \frac{p}{\eta} \left[(1 - \tau_k) \sum_{i \in N_1} C_{ik} |\mathbf{x}_i^T \boldsymbol{\beta} + b_{\tau_k} - y_i|^{p-1} \right. \\ & \left. - \tau_k \sum_{i \in N_2} C_{ik} |y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k}|^{p-1} \right], \end{aligned}$$

where $N_1 = \{i : y_i < \mathbf{x}_i^T \boldsymbol{\beta} + b_{\tau_k}\}$ and $N_2 = \{i : y_i > \mathbf{x}_i^T \boldsymbol{\beta} + b_{\tau_k}\}$. The acceptance probability is then given by

$$\varphi_{\mathbf{b}}(\mathbf{b}', \mathbf{b}) = 1 \wedge \frac{\pi(\mathbf{b}' | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{C}, \mathbf{w}, \eta, p) q(\mathbf{b} | \mathbf{b}')}{\pi(\mathbf{b} | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{C}, \mathbf{w}, \eta, p) q(\mathbf{b}' | \mathbf{b})},$$

where $q(\mathbf{b} | \mathbf{b}') = \mathcal{N}(\mathbf{b} | \mathbf{b}' + \epsilon_{\mathbf{b}}^2 / 2 \nabla_{\mathbf{b}'} \mathcal{L}(\mathbf{b}'), \epsilon_{\mathbf{b}}^2 \mathbf{I})$.

The conditional distribution of η is

$$\pi(\eta | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{C}, \mathbf{w}, p) \propto \frac{1}{\eta^{\frac{n}{p} + a_1}}$$

$$\exp \left\{ -\frac{1}{\eta} \left(\sum_{i=1}^n \sum_{k=1}^K C_{ik} \rho_{\tau_k, p} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k} \right) + a_2 \right) \right\}.$$

Hence

$$\eta | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{C}, \mathbf{w}, p \sim \mathcal{IG} \left(\frac{n}{p} + 1 + a_1, \sum_{i=1}^n \sum_{k=1}^K C_{ik} \rho_{\tau_k, p} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k} \right) + a_2 \right).$$

The conditional distribution of p is given by

$$\begin{aligned} \pi(p | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{C}, \mathbf{w}, \eta) &\propto \frac{1}{\Gamma(1 + 1/p)^n} \prod_{k=1}^K \left(\frac{1}{\tau_k^{\frac{1}{p}} + (1 - \tau_k)^{\frac{1}{p}}} \right)^{n_k} \\ &\exp \left\{ -\frac{1}{\eta} \sum_{i=1}^n \sum_{k=1}^K C_{ik} \rho_{\tau_k, p} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k} \right) \right\} I_{(0,5]}(p), \end{aligned} \tag{12}$$

where $I_{\mathcal{X}}(x)$ is the indicator function defined as 1 if $x \in \mathcal{X}$ and 0 otherwise. The normalizing constant of (12) is unknown, so p is sampled via Metropolis Hastings with proposals generated as $p' \sim \mathcal{N}_{(0,5]}(p, \epsilon_p^2)$, where $\mathcal{N}_{\mathcal{A}}(\mu, \sigma)$ denotes the normal distribution with location μ and variance σ truncated to the set \mathcal{A} . The corresponding acceptance probability is

$$\varphi_p(p', p) = 1 \wedge \frac{\pi(p' | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{C}, \mathbf{w}, \eta)}{\pi(p | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{C}, \mathbf{w}, \eta)} \frac{\Phi\left(\frac{5-p}{\epsilon_p}\right) - \Phi\left(\frac{-p}{\epsilon_p}\right)}{\Phi\left(\frac{5-p'}{\epsilon_p}\right) - \Phi\left(\frac{-p'}{\epsilon_p}\right)},$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

The conditional distribution of the component weights is

$$\begin{aligned} p(\mathbf{w} | \mathbf{y}, \mathbf{X}, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{C}, p, \eta) &\propto \prod_{k=1}^K w_k^{n_k + \alpha_k} \propto \text{Dirichlet}(n_1 + \alpha_1, \dots, n_K + \alpha_K). \end{aligned}$$

The conditional distribution of $\mathbf{C}_i = (C_{i1}, \dots, C_{iK})^T$ is a multinomial distribution

$$\begin{aligned} p(\mathbf{C}_i | y_i, \mathbf{x}_i, \boldsymbol{\tau}, \boldsymbol{\beta}, \mathbf{b}, p, \eta) &\propto \prod_{k=1}^K \left[\frac{w_k}{\tau_k^{\frac{1}{p}} + (1 - \tau_k)^{\frac{1}{p}}} \exp \left\{ -\eta^{-1} \rho_{\tau_k, p} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta} - b_{\tau_k} \right) \right\} \right]^{C_{ik}} \end{aligned}$$

$$\propto \text{Multinomial}(1, \hat{p}_1, \dots, \hat{p}_k),$$

where

$$\hat{p}_k = \frac{w_k [\tau_k^{-1/p} + (1 - \tau_k)^{-1/p}] \exp\{-C_{ik} \rho_{\tau_k, p}(y_i - b_{\tau_k} - \mathbf{x}_i^T \boldsymbol{\beta}) / \eta\}}{\sum_{k=1}^K w_k [\tau_k^{-1/p} + (1 - \tau_k)^{-1/p}] \exp\{-C_{ik} \rho_{\tau_k, p}(y_i - b_{\tau_k} - \mathbf{x}_i^T \boldsymbol{\beta}) / \eta\}}.$$

3 Numerical studies

3.1 Simulation

In this section, Monte Carlo simulations are done to compare the performance of Bayesian regularized composite L^p -quantile regression and Bayesian regularized composite quantile regression. Julia (Bezanson et al. 2017) was used to produce the results.¹ Note that priors for BCQR are set as in Huang and Chen (2015). We consider the linear model where data are generated from

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n, \quad (13)$$

where \mathbf{x}_i is sampled from $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and multiple error distributions are considered for ϵ_i . Note that the error distributions have been chosen based on those considered in Huang and Chen (2015), and all results are stable over different parameters than those reported. In the sequel, MN shall denote the mixture of normal distributions, $0.5\mathcal{N}(-2, 1) + 0.5\mathcal{N}(2, 1)$, and ML the mixture of Laplace distributions, $0.5\text{Lap}(-2, 1) + 0.5\text{Lap}(2, 1)$ and the dimension of $\boldsymbol{\beta}$ is denoted by m . Two settings are considered where the first is the dense case with $n = 200$ and $m = 8$ where $\boldsymbol{\beta} = \mathbf{1}$. For the second setting, the sparse case, $n = 100$ and $m = 20$ with $(\beta_1, \beta_2, \beta_5)^T = (0.5, 1.5, 0.2)^T$, where β_j is the j th position of $\boldsymbol{\beta}$, and the remaining coefficients are set to 0. The root mean square error (RMSE) is used to compare different methods, $\text{RMSE}(\hat{\boldsymbol{\beta}}) = E(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|)$, where $\hat{\boldsymbol{\beta}}$ is taken as the mean of the posterior sample. For each simulated datum, the first 3000 sweeps of the chains are discarded as burn-in. Then, an additional 10,000 sweeps are performed, with every 5th sweep kept to reduce the serial correlation of the chains. As in Huang and Chen (2015), the number of components is fixed to $K = 9$ with $\tau_k = k/(K + 1)$ for $k = 1, \dots, 9$. Other values of K have been considered, however, results are not sensitive to this choice, so results over multiple K are not displayed. The simulations are repeated 1000 times.

Results in terms of RMSE for both settings are found in Table 1. The proposed method outperforms BCQR for all distributions except the mixture Laplace distribution for setting 1, where the results are very similar. As expected, the results are close for the Laplace error distribution, highlighting that BCQR is a special case of BCLQR, with $p = 1$.

For the sparse case, the variable selection properties of BCQR and BCLQR are compared. Denote the number of correctly classified non-zero coefficients, true posi-

¹ Version 1.9.0 was used, code can be found at <https://github.com/lukketotte/CompositeLPQR>.

Table 1 Summary table of RMSE over 1000 replications of simulated data, standard errors in parenthesis

		$\mathcal{N}(0, 1)$	MN	Lap(0, 1)	ML	Gam(1, 1)	$\mathcal{B}(2, 2)$
Setting 1	BCQR	0.215 (0.055)	0.265 (0.078)	0.228 (0.068)	0.282 (0.092)	0.066 (0.022)	0.042 (0.011)
	BCLQR	0.199 (0.051)	0.246 (0.066)	0.224 (0.065)	0.284 (0.099)	0.057 (0.021)	0.037 (0.009)
Setting 2	BCQR	0.495 (0.096)	0.725 (0.205)	0.590 (0.136)	0.879 (0.253)	0.273 (0.069)	0.110 (0.020)
	BCLQR	0.464 (0.085)	0.683 (0.180)	0.584 (0.126)	0.876 (0.211)	0.269 (0.072)	0.103 (0.022)

Table 2 Summary of variable selection by comparison of true and false positives and overall accuracy for the sparse setting with mean over 1000 replications of simulated data reported, standard errors in parenthesis

		$\mathcal{N}(0, 1)$	MN	Lap(0, 1)	ML	Gam(1, 1)	$\mathcal{B}(2, 2)$
TP	BCQR	2.422	1.996	2.214	1.721	2.837	3.0
	BCLQR	2.439	1.936	2.214	1.617	2.860	3.0
FP	BCQR	0.578	1.004	0.787	1.279	0.163	0.0
	BCLQR	0.561	1.064	0.787	1.383	0.140	0.0
OA	BCQR	0.921 (0.075)	0.898 (0.077)	0.916 (0.063)	0.893 (0.077)	0.953 (0.068)	0.952 (0.068)
	BCLQR	0.935 (0.059)	916 (0.058)	0.920 (0.054)	0.901 (0.056)	0.960 (0.068)	0.970 (0.04)

tives, by TP and the number of incorrectly classified zero coefficients, false positives, by FP. A coefficient is classified as non-zero if the 95% highest posterior density interval does not cover zero. Note that both TP and FP are at most 3 from the specification of β . Otherwise, it is classified as zero. We also denote a correctly classified 0 as a true negative (TN), and define overall accuracy (OA) as $(TN + TP)/20$. Table 2 shows that the methods perform similarly, in terms of TP and FP, except for the mixture Laplace case, where BCQR performs better than BCLQR. Further, for the OA, the proposed method outperforms BCQR with respect to all error distributions.

3.2 Empirical data

In this section, the proposed method is applied to empirical data and compared with BCQR. As in Huang and Chen (2015), 10-fold cross-validation is used to evaluate the performance of the two methods, with accuracy measured by the mean absolute prediction error (MAPE) and corresponding standard deviation evaluated on the test data. As in the simulation study, the number of components is fixed to $K = 9$ with $\tau_k = k/(K + 1)$ for $k = 1, \dots, 9$.

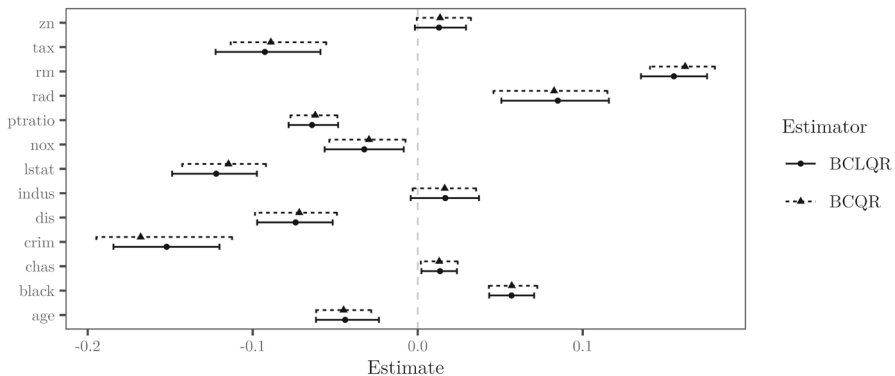


Fig. 1 MAP estimates and 95% credible sets for all parameters based on BCLQR and BCQR for the entire Boston housing data

3.2.1 Boston housing data

The Boston housing data were first analyzed by Harrison Jr and Rubinfeld (1978) and has been used extensively in the context of Bayesian quantile regression, see, e.g., Huang and Chen (2015); Kozumi and Kobayashi (2011); Li et al. (2010). The data consists of 506 observations and were obtained from the MASS (Venables and Ripley 2002) package in R (R Core Team 2021).² The relationship between log-transformed median value of owner-occupied housing (in 1000 USD) and the remaining 13 variables is considered, for details on the parameters see Venables and Ripley (2002). We draw 30,000 samples from the posterior with the first 5000 discarded as the transient phase. The posterior sample is thinned down to every 5th sample.

The MAPE of BCLQR is 0.137, with a standard deviation of 0.032. Similarly, the MAPE of BCQR is 0.139, with a standard deviation of 0.031. Thus we find that the proposed method performs slightly better. In Fig. 1, the estimators are compared for the complete data, where differences are generally negligible, however the parameter *Crim* stands out in terms of interval width and MAP estimate. The intervals of Fig. 1 shows that the the procedures gives the same result in terms of variable selection. In Table 3, MAP estimates and corresponding standard deviation for six regression coefficients are presented. Table 3 also includes effective sample size (ESS), which shows that the efficiency in exploration of the posterior for the two MCMC sampling schemes are quite different. BCLQR is generally more efficient. For more details on ESS, see, e.g., Section 11.5 of Gelman et al. (2014).

In Fig. 2, the result for quantile specific intercepts and mixture weights are displayed. The main difference between BCLQR and BCQR is that the weights of the former concentrates to a much larger degree than the latter, as seen in Fig. 2b. The concentration explains the larger degree of oscillation of the intercepts for BCLQR over quantiles, in relation to the intercepts of BCQR, as seen in Fig. 2a.

² R version used was 4.3.0 and MASS package version 7.3.

Table 3 ESS of the MCMC procedures and estimated regression coefficients for the entire Boston housing data, with standard error in parenthesis

Parameter	Estimate		ESS	
	BCLQR	BCQR	BCLQR	BCQR
Zn	0.012 (0.009)	0.013 (0.009)	795	483
Tax	-0.091 (0.018)	-0.087 (0.016)	1216	141
Rm	0.149 (0.011)	0.164 (0.013)	39	210
Rad	0.085 (0.017)	0.082 (0.018)	1105	138
Ptatio	-0.061 (0.008)	-0.062 (0.007)	1005	473
Nox	-0.030 (0.013)	-0.029 (0.012)	256	261
Lstat	-0.130 (0.014)	-0.130 (0.014)	61	231
Indus	0.017 (0.012)	0.018 (0.011)	1336	283
Dis	-0.071 (0.011)	-0.070 (0.012)	855	262
Crim	-0.171 (0.140)	-0.175 (0.202)	140	68
Chas	0.016 (0.006)	0.014 (0.005)	1006	1156
Black	0.055 (0.007)	0.056 (0.007)	836	569
Age	-0.043 (0.013)	-0.045 (0.010)	589	209

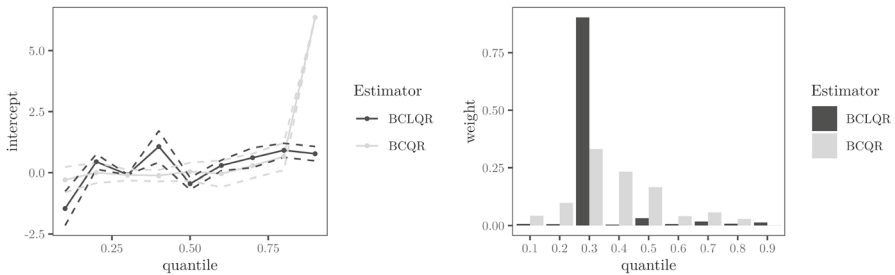


Fig. 2 Comparison over quantiles $\tau \in \{0.1, 0.2, \dots, 0.9\}$ for estimation on entire Boston data

3.2.2 Body measurement data

In this section, the body measurement data, which was first analysed by Heinz et al. (2003), is considered. The data consists of 507 observations and were downloaded from the `Brq` package in R (R Core Team 2021).³ The relationship between body weight and 24 other variables is considered. See Heinz et al. (2003) for details on the data and descriptions of the independent variables. We draw 30,000 samples from the posterior with the first 5000 discarded as the transient phase. The posterior sample is thinned down to every 5th sample.

The MAPE of BCLQR is 1.577, with a standard deviation of 0.221. The MAPE of BCQR is 1.680 with a standard deviation of 0.242. Thus we find that the proposed

³ `Brq` package version 3.0.

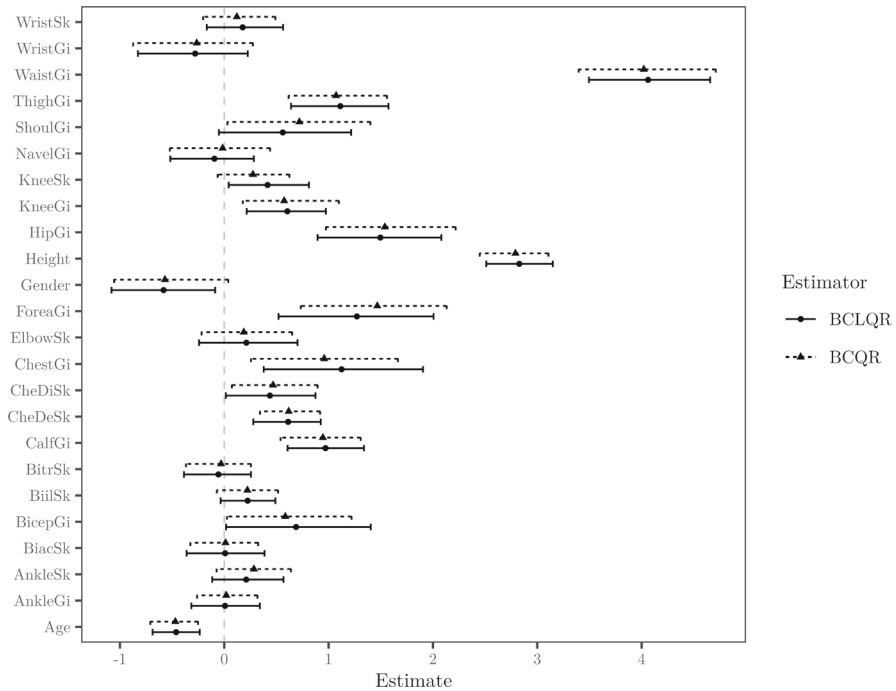


Fig. 3 Point estimates and 95% credible sets for all parameters based on BCLQR and BCQR for the entire Body measurement data

method performs better than BCQR. The result from the estimation of the whole sample is shown in Fig. 3, with 95% credible sets. In terms of variable selection, the procedures differ on the variables *ShouldGi*, *KneeSk* and *Gender*. The interval of BCLQR for *ShouldGi* covers 0 whilst that of BCQR does not, vice versa for *KneeSk* and *Gender*.

MAP estimates, standard deviation and ESS for all regression coefficients are presented in Table 4. The difference in ESS for the posterior samples of β is much greater for the Body measurement data in comparison to the Boston housing data. This is most likely due to the joint sampling of β in our MCMC procedure, as compared to each index of β being sampled individually in BCQR (Huang and Chen 2015).

In Fig. 4, the results for quantile specific intercepts and mixture weights are displayed where patterns similar to those noted in Fig. 2 can be seen.

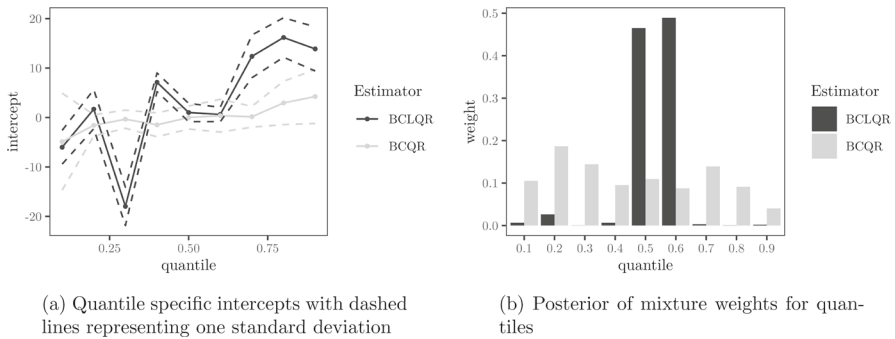


Fig. 4 Comparison over quantiles $\tau \in \{0.1, 0.2, \dots, 0.9\}$ for entire Body measurement data

Table 4 ESS of the MCMC procedures and estimated regression coefficients for the entire body measurement data, with standard error in parenthesis

Parameter	Estimate		ESS	
	BCLQR	BCQR	BCLQR	BCQR
WristSk	0.187 (0.184)	0.131 (0.181)	1813	80
WristGi	-0.284 (0.267)	-0.288 (0.297)	777	39
WaistGi	4.063 (0.291)	4.035 (0.326)	1498	26
ThighGi	1.113 (0.237)	1.090 (0.237)	1071	50
ShoulGi	0.569 (0.324)	0.710 (0.353)	311	45
NavelGi	-0.099 (0.205)	-0.027 (0.250)	1591	46
KneeSk	0.420 (0.189)	0.276 (0.186)	896	56
KneeGi	0.604 (0.197)	0.591 (0.237)	1271	70
HipGi	1.493 (0.303)	1.551 (0.328)	1211	49
Height	2.83 (0.165)	2.786 (0.163)	1167	91
Gender	-0.581 (0.252)	-0.559 (0.281)	1283	39
ForeaGi	1.266 (0.373)	1.471 (0.367)	905	31
ElbowSk	0.218 (0.226)	0.193 (0.221)	939	62
ChestGi	1.131 (0.387)	0.947 (0.369)	554	51
CheDiSk	0.439 (0.219)	0.471 (0.213)	750	85
CheDeSk	0.609 (0.161)	0.618 (0.145)	1203	168
CalfGi	0.972 (0.184)	0.942 (0.193)	1203	109
BitrSk	-0.060 (0.159)	-0.037 (0.151)	1677	104
BiilSk	0.223 (0.134)	0.221 (0.150)	1072	100
BicepGi	0.691 (0.343)	0.611 (0.320)	961	35
BiacSk	0.009 (0.182)	0.006 (0.164)	831	120
AnkleSk	0.216 (0.175)	0.287 (0.175)	1261	93
AnkleGi	0.010 (0.160)	0.021 (0.146)	1937	164
Age	-0.460 (0.115)	-0.472 (0.118)	1048	159

4 Conclusions

In this paper, we present an efficient Bayesian method that extends the approach of Huang and Chen (2015) to combine multiple L^p -quantile regressions for both inference and variable selection. A simulation study demonstrates that the proposed method performs as well as, or even better than, Bayesian composite quantile regression (BCQR). BCQR outperforms the proposed method only in terms of true positives for variable selection accuracy for some specific error distributions. However, when considering true negatives as well, the proposed method surpasses BCQR across all error distributions. This performance improvement holds true for both sparse and dense settings, underscoring the versatility of L^p -quantiles, which encompass quantiles and expectiles through the parameter p as special cases, in contrast to ordinary quantiles.

Comparisons using empirical data reveal that the proposed method outperforms BCQR in terms of prediction error during 10-fold cross-validation. Furthermore, the overall effective sample size of the regression coefficients in Bayesian composite L^p -quantile regression (BCLQR) is higher for the parameters of interest in practical applications, indicating more efficient exploration of the posterior distribution.

For future research, it may be valuable to relax the assumption of a fixed number of quantiles in the composite model. Additionally, treating discretized quantiles, $\{\tau_k\}_{k=1}^K$, as continuous variables and considering a Dirichlet process as a prior on the component weights could be explored. Extending BCQR to infinite mixtures remains an open and intriguing possibility. Another promising avenue is to study the specific case of BCLQR with a fixed $p = 2$, essentially composite expectile regression.

Acknowledgements The authors are thankful to the editor and the reviewers for their help and suggestions for the improvement of the article.

Funding Open access funding provided by Uppsala University.

Declarations

Conflicts of interest The author declares that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arnroth L, Vegelius J (2023) Quantile regression based on the skewed exponential power distribution. *Commun Stat Simul Comput*
- Artzner P, Delbaen F, Eber J-M, Heath D (1999) Coherent measures of risk. *Math Financ* 9(3):203–228
- Bellini F, Klar B, Müller A, Gianin ER (2014) Generalized quantiles as risk measures. *Insur Math Econ* 54:41–48

- Bernardi M, Bottone M, Petrella L (2018) Bayesian quantile regression using the skew exponential power distribution. *Comput Stat Data Anal* 126:92–111
- Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: a fresh approach to numerical computing. *SIAM Rev* 59(1):65–98. <https://doi.org/10.1137/141000671>
- Chen Z (1996) Conditional L_p -quantiles and their application to the testing of symmetry in non-parametric regression. *Stat. Probab. Lett.* 29(2):107–115
- Daouia A, Girard S, Stupfler G (2019) Extreme m -quantiles as risk measures: from l^1 to l^p optimization. *Bernoulli* 25(1):264–309
- Gelman A, Carlin JB, Stern HS, Rubin DB (2014) Bayesian data analysis, 3rd edn. Chapman and Hall/CRC, New York
- Harrison D Jr, Rubinfeld DL (1978) Hedonic housing prices and the demand for clean air. *J Environ Econ Manag* 5(1):81–102
- Heinz G, Peterson LJ, Johnson RW, Kerk CJ (2003) Exploring relationships in body dimensions. *J Stat Educ* 11(2)
- Huang H, Chen Z (2015) Bayesian composite quantile regression. *J Stat Comput Simul* 85(18):3744–3754
- Jones MC (1994) Expectiles and m -quantiles are quantiles. *Stat Probab Lett* 20(2):149–153
- Koenker R, Bassett G (1978) Regression quantiles. *Econometrica* 46(1):33–50
- Komunjer I (2007) Assymmetric power distribution: theory and applications to risk measurement. *J Appl Econom* 22(5):891–921
- Konen D, Paindaveine D (2022) Multivariate ρ -quantiles: a spatial approach. *Bernoulli* 28(3):1912–1934
- Kozumi H, Kobayashi G (2011) Gibbs sampling methods for Bayesian quantile regression. *J Stat Comput Simul* 81(11):1565–1578
- Li Q, Lin N, Xi R (2010) Bayesian regularized quantile regression. *Bayesian Anal* 5(3):533–556
- Newey WK, Powell JL (1987) Asymmetric least squares estimation and testing. *Econometrica* 55(4):819–847
- R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Stat Methodol* 58(1):267–288
- Usseglio-Carleve A (2018) Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. *Electron J Stat* 12(2):4057–4093
- Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York
- Zhao Z, Xiao Z (2014) Efficient regressions via optimally combining quantile information. *Econom Theory* 30(6):1272–1314
- Zhu D, Zinde-Walsh V (2009) Properties and estimation of asymmetric exponential power distribution. *J Econom* 148(1):86–99
- Zou H, Yuan M (2008) Composite quantile regression and the oracle model selection theory. *Ann Stat* 36(3):1108–1126