



Bayesian additive tree ensembles for composite quantile regressions

Yaeji Lim¹ · Ruijin Lu² · Madeleine St. Ville³ · Zhen Chen³

Received: 1 May 2025 / Accepted: 14 August 2025

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2025

Abstract

In this paper, we introduce a novel approach that integrates Bayesian additive regression trees (BART) with the composite quantile regression (CQR) framework, creating a robust method for modeling complex relationships between predictors and outcomes under various error distributions. Unlike traditional quantile regression, which focuses on specific quantile levels, our proposed method, composite quantile BART, offers greater flexibility in capturing the entire conditional distribution of the response variable. By leveraging the strengths of BART and CQR, the proposed method provides enhanced predictive performance, especially in the presence of heavy-tailed errors and non-linear covariate effects. Numerical studies confirm that the proposed composite quantile BART method generally outperforms classical BART, quantile BART, and composite quantile linear regression models in terms of RMSE, especially under heavy-tailed or contaminated error distributions. Notably, under contaminated normal errors, it reduces RMSE by approximately 17% compared to composite quantile regression, and by 27% compared to classical BART.

Keywords Bayesian additive regression trees · Composite quantile regression · Heavy-tailed errors · Non-linear covariate effects

1 Introduction

Quantile methods are popular approaches for handling non-Gaussian data in regression modeling framework (Koenker and Hallock 2001; Hao and Naiman 2007). Unlike ordinary least squares (OLS) regression that focuses on the mean outcome given predictor variables, quantile regression (QR) examines the effects of covariates on the entire distribution of the response variable, offering a more comprehensive char-

acterization of the data. This method provides robust results in the presence of heavy-tailed errors or outliers (Koenker 2005), making it a valuable tool for many applied settings. QR models have been successfully applied in epidemiology (Lee and Neocleous 2010; Wei et al. 2019), climatology (Haugen et al. 2018; Reich et al. 2011), and economics (Fitzenberger et al. 2013; Marrocu et al. 2015).

Except in situations where a particular quantile level is of interest (e.g., growth percentiles (Wei et al. 2006; Chen and Müller 2012)), the choice of appropriate quantile levels in quantile analysis affects the relative efficiency of the estimators, presenting a challenge in practical applications (Koenker and Bassett Jr 1978; Zhao and Xiao 2014). Zheng et al. (2015) also highlighted challenges in traditional quantile regression approaches that focus on estimating covariate effects at a single or a few prespecified quantile levels, noting the lack of a clear scientific basis for choosing one quantile over a nearby alternative. Additionally, it has been shown that QR can have arbitrarily low relative efficiency compared to OLS. To overcome the drawbacks of traditional QR, Zou and Yuan (2008) proposed a composite quantile regression (CQR) method to address multiple quantile regression models concurrently. Since then, significant efforts have been made to extend CQR. Jiang et al. (2012) pointed out that applying the same weight across different quantile lev-

✉ Zhen Chen
chenzhe@mail.nih.gov

Yaeji Lim
yaeji.lim@gmail.com

Ruijin Lu
r.lu@wustl.edu

Madeleine St. Ville
maddy.stville@nih.gov

¹ Department of Applied Statistics, Chung-Ang University, Seoul, Korea

² Institute for Informatics, Data Science & Biostatistics, Washington University in St. Louis, Missouri, USA

³ Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Maryland, USA

els is typically suboptimal and introduced weighted CQR (WCQR), which was later enhanced by Zhao and Lian (2016) to improve its efficiency. Unlike traditional QR, CQR does not require selecting specific quantiles and retains the robustness and other desirable properties of the quantile method (Zhao and Xiao 2014). In addition, when the error variance is finite, CQR still enjoys great advantages in terms of estimation efficiency. This approach enhances the flexibility and efficiency of the QR framework, allowing for a more holistic analysis of the response variable's distribution. Huang and Chen (2015), Xu et al. (2017), and Yuan et al. (2023) are a few examples where CQR have been shown to outperform regular QR.

Both frequentist and Bayesian approaches to QR and CQR are abundant. Under the frequentist framework, Taylor (2000) proposed a quantile regression neural network (QRNN) to estimate the conditional probability distribution of multiperiod financial returns, and Jiang et al. (2013) extended a CQR estimation procedure for single index models. Galvao and Kato (2016) and Powell (2022) developed QR models for panel data. Bayesian approaches to QR make use of the equivalence between the minimization of the loss with the quantile check function of Koenker (2005) and maximization of likelihood function with an asymmetric Laplace distribution (ALD) error term (Yu and Moyeed 2001; Sriram et al. 2013). Using the mixture representation proposed by Kozumi and Kobayashi (2011), a Gibbs sampler for Bayesian QR can be implemented. This approach facilitates the estimation of complex models and the incorporation of prior information, enhancing the overall inferential process. For example, Li et al. (2010) explored regularization in Bayesian QR, while Yang et al. (2016) established the asymptotic validity of posterior inference for pseudo-Bayesian QR methods using an asymmetric Laplace likelihood. Additionally, Benoit and Van den Poel (2017) developed an R package for estimating QR parameters through a Bayesian approach based on the asymmetric Laplace distribution. These developments illustrate the potential of Bayesian methods to address various challenges in QR, including model complexity and computational efficiency. For the composite quantile model, Huang and Chen (2015) proposed a WCQR model where the weight of each component can be treated as an unknown parameter and estimated via Markov chain Monte Carlo (MCMC) sampling in a Bayesian hierarchical framework. This approach allows for greater flexibility in modeling and can lead to improved predictive performance by optimally combining information from multiple quantiles.

In many applications, the linearity assumption between covariates and conditional quantiles might not hold. In these situations, semi- or nonparametric approaches are attractive alternatives. For example, Koenker et al. (1994) considered regression spline approaches for estimating the conditional QR and Yu and Jones (1998) proposed local linear poly-

mial QR. Similarly, Kai et al. (2010) developed the local polynomial CQR estimators and proved its efficiency for non-normal error distributions. For Bayesian approaches, Thompson et al. (2010) proposed a nonparametric Bayesian QR method using natural cubic splines, offering a flexible alternative to parametric Bayesian QR models, particularly when the linearity assumption fails. More recently, Xu and Reich (2023) proposed a nonlinear simultaneous QR model by specifying a Bayesian nonparametric model for the conditional distribution.

Recently, Bayesian regression trees and their ensembles have demonstrated enhanced predictive performance in least squares regression, binary classification, and multiclass classification contexts. These methods have garnered significant attention due to their flexibility and ability to model complex relationships between variables. Notably, Bayesian additive regression trees (BART) (Chipman et al. 2012) estimate the conditional mean of a response given a set of predictors using a sum of regression trees model, showing remarkable predictive performance across various applications (Sparapani et al. 2016; Zhang et al. 2020). This approach leverages the power of multiple regression trees to capture non-linear interactions and intricate patterns in data, making it a robust tool for various statistical modeling tasks. Furthermore, Linero (2018) demonstrated the utility of the Dirichlet splitting probability prior within the BART framework for both prediction and variable selection problems. Additionally, Linero and Yang (2018) introduced soft decision trees and sparsity-inducing priors in BART, illustrating their promising performance. Several extensions of BART have been proposed to handle different types of outcomes. Notably, Kindo (2016) developed BART-based methods for multinomial, ordinal, and quantile regression as part of his dissertation. Subsequent works further advanced these directions, including multinomial probit BART (Xu et al. 2024) and BART for ordinal outcomes (Lee and Hwang 2024). Basak et al. (2022) developed BART for censored survival data, and Um et al. (2023) extended BART to multivariate skewed responses. These extensions have broadened the applicability of BART, enabling it to address a wider range of statistical challenges beyond traditional regression. Although BART has primarily been used for mean regression, several recent studies have extended it to quantile or tail estimation. For example, Clark et al. (2023) used BART-based vector autoregressions to perform real-time tail forecasting of GDP growth, inflation, and unemployment. In addition, Clark et al. (2024) employed a nonparametric quantile panel regression model that uses BART to capture nonlinear effects, and Baumeister et al. (2024) proposed a mixture BART model with stochastic volatility for forecasting the tails of the conditional distribution. In contrast to these applications, Kindo et al. (2016) developed a quantile version of BART (QBART) by incorporating the asymmetric Laplace distribution into the model,

allowing for direct estimation of conditional quantiles. They demonstrated its superiority over linear quantile regression and quantile random forests.

In this paper, we consider the integration of BART into the composite quantile method, creating a BART for composite quantile regression (BART-CQR) to improve prediction accuracy under a wide range of error distributions. Rather than targeting a specific conditional quantile, BART-CQR aims to produce robust and efficient estimates of the conditional mean by aggregating information across multiple quantile levels. This aligns with the original motivation of CQR (Zou and Yuan 2008), which seeks robustness against heavy-tailed and non-Gaussian errors while maintaining estimation efficiency. Related work by Cao et al. (2024) proposed an adaptive trimmed regression approach based on BART, which enhances robustness by incorporating data-driven tuning parameters. However, their method focuses on effectively identifying suspected outliers and removing them from the analysis. In contrast, our BART-CQR method takes a different approach to robustness: rather than detecting and trimming outliers, it leverages the composite quantile framework, which inherently provides robustness to distributional misspecification without excluding any observations. This design enables BART-CQR to utilize the full dataset while maintaining efficiency and robustness in estimating the conditional mean.

Our approach extends the work of QBART by Kindo et al. (2016), which focused on single quantile regression, by generalizing it to composite quantile regression. Therefore, our focus is on accurate and robust prediction of the conditional mean, not on modeling specific quantiles. Our method also builds upon existing Bayesian composite quantile methods (Huang and Chen 2015; Alhamzawi 2016), which are based on linear models, by introducing a flexible additive regression tree framework to better capture nonlinear relationships between outcomes and predictors. This model is a fully Bayesian framework for constructing composite quantile regression trees and their ensembles. Through numerical studies, we verify that the BART-CQR outperforms the classical BART, QBART and CQR models under heavy-tailed distributions and in the presence of non-linearity between the input and output variables. This indicates that the proposed method combines the benefits of ensemble trees and the composite quantile approach, offering a powerful tool for handling complex and heavy-tailed data distributions.

The rest of the article is organized as follows. Section 2 introduces the BART-CQR method in detail, explaining its underlying principles and key features. Section 3 provides details on the posterior sampling of BART-CQR. Section 4 presents simulation results, and Section 5 illustrates a real data application. Conclusion and discussion are presented in Section 6. The R codes for implementing the numerical

experiments in this study are available at <https://github.com/yaeji-lim/BARTCQR>.

2 BART for composite quantile regressions

2.1 Bayesian QR and CQR

Given $(y_i, \mathbf{x}_i)$ for $i = 1, \dots, n$, where $y_i \in \mathbb{R}$ and $\mathbf{x}_i \in \mathbb{R}^p$, consider the following regression model:

$$y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad (1)$$

where β_0 is an intercept, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the vector of unknown coefficients and $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{ALD}(\tau, \theta)$. The density function of $\text{ALD}(\tau, \theta)$ is $f_\theta(x|\tau) = \theta(1-\theta)\tau e^{-\tau\rho_\theta(x)}$, where $\theta \in (0, 1)$ is an asymmetry parameter, τ is a precision parameter, and $\rho_\theta(t) = t(\theta - \mathbf{I}_{(t < 0)})$ is the check loss function. Then, the conditional θ th quantile of $y_i|\mathbf{x}_i$ is

$$\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + q_{\epsilon_i} := b_\theta + \mathbf{x}_i^T \boldsymbol{\beta},$$

where q_{ϵ_i} is the θ th quantile of ϵ_i , and QR estimates coefficients by solving following minimization:

$$(\hat{b}_\theta, \hat{\boldsymbol{\beta}}) = \arg \min_{b_\theta, \boldsymbol{\beta}} \sum_{i=1}^n \rho_\theta(y_i - b_\theta - \mathbf{x}_i^T \boldsymbol{\beta}). \quad (2)$$

The minimization is exactly equivalent to the maximization of a likelihood function,

$$\prod_{i=1}^n \left\{ \theta(1-\theta)\tau \exp \left[-\tau\rho_\theta(y_i - b_\theta - \mathbf{x}_i^T \boldsymbol{\beta}) \right] \right\}. \quad (3)$$

The minimum criterion (2) can be extended to the weighted CQR with multiple quantiles $\theta_1, \dots, \theta_K$ as follows:

$$\arg \min_{(b_{\theta_1}, \dots, b_{\theta_K}), \boldsymbol{\beta}} \sum_{i=1}^n \sum_{k=1}^K \omega_k \rho_{\theta_k}(y_i - b_{\theta_k} - \mathbf{x}_i^T \boldsymbol{\beta}),$$

where $0 \leq \omega_k \leq 1$ is the weight for the k th component with $\sum_k \omega_k = 1$. We extend the joint distribution of (3) to the composite model as:

$$\prod_{i=1}^n \left\{ \sum_{k=1}^K \omega_k \theta_k (1-\theta_k) \tau \exp \left[-\tau\rho_{\theta_k}(y_i - b_{\theta_k} - \mathbf{x}_i^T \boldsymbol{\beta}) \right] \right\}.$$

Due to the complexity of directly solving this, it is common to introduce a cluster matrix $\mathbf{C}$, where the (i, k) th element, the latent variable C_{ik} , is equal to 1 if the i th subject belongs to the k th cluster; otherwise, $C_{ik} = 0$. We assume that each

observation belongs to exactly one cluster, i.e., for each i , $\sum_{k=1}^K C_{ik} = 1$. The complete likelihood is then:

$$\begin{aligned} & \prod_{i=1}^n [P(y_i | \mathbf{C}_i) \times P(\mathbf{C}_i)] \\ &= \prod_{i=1}^n \prod_{k=1}^K \left\{ \omega_k \theta_k (1 - \theta_k) \exp \left[-\tau \rho_{\theta_k} (y_i - b_{\theta_k} - \mathbf{x}_i^T \boldsymbol{\beta}) \right] \right\}^{C_{ik}}, \end{aligned} \quad (4)$$

where $\mathbf{C}_i = (C_{i1}, \dots, C_{iK})$.

We place a Laplace prior on $\boldsymbol{\beta}$ for regularization:

$$\pi(\boldsymbol{\beta} | \tau, \lambda) = \left(\frac{\tau \lambda}{2} \right)^p \exp \left(-\tau \lambda \sum_{j=1}^p |\beta_j| \right),$$

and the prior can be further represented as

$$\begin{aligned} \pi(\boldsymbol{\beta} | \eta^2) &= \prod_{j=1}^p \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} \exp \left(-\frac{\beta_j^2}{2s_j} \right) \frac{\eta^2}{2} \\ &\quad \times \exp \left(-\frac{\eta^2}{2} s_j \right) ds_j, \end{aligned}$$

where $\eta := \tau \lambda$. For $\pi(\boldsymbol{\omega})$, we assume $\pi(\boldsymbol{\omega}) = \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$ with $\boldsymbol{\omega} = (\omega_1, \dots, \omega_K)^T$. The priors for τ and η^2 are assumed to follow gamma distributions.

Then the posterior distribution is given by:

$$\begin{aligned} & \prod_{i=1}^n \prod_{k=1}^K \left\{ \omega_k \theta_k (1 - \theta_k) \exp \left[-\tau \rho_{\theta_k} (y_i - b_{\theta_k} - \mathbf{x}_i^T \boldsymbol{\beta}) \right] \right\}^{C_{ik}} \\ & \times \pi(\boldsymbol{\beta} | \eta^2) \pi(\tau, \eta^2) \pi(\boldsymbol{\omega}). \end{aligned}$$

To obtain closed-form conditional distributions, we use the representation of ALD as a mixture of an exponential and a scaled normal distribution (Kozumi and Kobayashi 2011). The regression model (1) can then be expressed as:

$$\begin{aligned} y_i &= \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \tau^{-1} \xi_1 v_i + \tau^{-1} \xi_2 \sqrt{v_i} z_i \\ &= \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \xi_1 \tilde{v}_i + \tau^{-1/2} \xi_2 \sqrt{\tilde{v}_i} z_i, \end{aligned}$$

where $\xi_1 = \frac{1-2\theta}{\theta(1-\theta)}$ and $\xi_2^2 = \frac{2}{\theta(1-\theta)}$. Here, $v_i \sim \exp(1)$, $\tilde{v}_i := \tau^{-1} v_i \sim \exp(\tau^{-1})$ and $z_i \sim N(0, 1)$ are independent. The hierarchical model for MCMC sampling is then:

$$\begin{aligned} y_i &= b_{\theta_k} + \mathbf{x}_i^T \boldsymbol{\beta} + \xi_{1k} \tilde{v}_i \\ &\quad + \tau^{-1/2} \xi_{2k} \sqrt{\tilde{v}_i} z_i, \quad \text{for all } i \text{ such that } C_{ik} = 1, \\ \tilde{\mathbf{v}} | \tau &\sim \prod_{i=1}^n \tau \exp(-\tau \tilde{v}_i) \quad \text{for } \tilde{\mathbf{v}} = (\tilde{v}_1, \dots, \tilde{v}_n)^T, \end{aligned}$$

$$\begin{aligned} \mathbf{z} &\sim \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} z_i^2 \right) \quad \text{for } \mathbf{z} = (z_1, \dots, z_n)^T, \\ \boldsymbol{\beta}, \mathbf{s} | \eta^2 &\sim \prod_{j=1}^p \frac{1}{\sqrt{2\pi s_j}} \exp \left(-\frac{\beta_j^2}{2s_j} \right) \prod_{j=1}^p \frac{\eta^2}{2} \exp \left(-\frac{\eta^2}{2} s_j \right) \\ &\quad \text{for } \mathbf{s} = (s_1, \dots, s_p)^T, \\ (\tau, \eta^2) &\sim \tau^{a_\tau - 1} \exp(-b_\tau \tau) (\eta^2)^{a_\eta - 1} \exp(-b_\eta \eta^2), \\ \boldsymbol{\omega} &\sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K), \end{aligned}$$

where $a_\tau, b_\tau, a_\eta, b_\eta$ are hyperparameters.

Denote $\mathbf{y} = (y_1, \dots, y_n)^T$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$, and $\mathbf{b} = (b_{\theta_1}, \dots, b_{\theta_K})^T$. The complete likelihood based on ALD form is:

$$\begin{aligned} f(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \mathbf{s}, \eta^2, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \boldsymbol{\omega}, \mathbf{C}) &= \\ & \prod_{i=1}^n \prod_{k=1}^K \left(\frac{1}{\sqrt{2\pi \tau^{-1} \xi_{2k}^2 \tilde{v}_i}} \right)^{C_{ik}} \\ & \exp \left[-\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^K \frac{C_{ik} (y_i - b_{\theta_k} - \mathbf{x}_i^T \boldsymbol{\beta} - \xi_{1k} \tilde{v}_i)^2}{\tau^{-1} \xi_{2k}^2 \tilde{v}_i} \right]. \end{aligned}$$

2.2 BART CQR

To allow a semiparametric relationship between outcome and predictors, consider the following regression model:

$$y_i = h(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (5)$$

where h is an unknown function, and $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{ALD}(\tau, \theta)$. The idea of applying BART to CQR is to model (5) as

$$y_i = b_{\theta_k} + \sum_j g(x_i; T_j, M_j) + \epsilon_i, \quad \text{for all } i \text{ such that } C_{ik} = 1,$$

where $(\mathbf{T}, \mathbf{M}) = (T_j, M_j); j = 1, \dots, n_T$, with T_j and M_j being the parameters of the j th tree in the BART model.

We assume that the priors on any two distinct trees in the sum are independent and the prior on τ is independent of the tree priors. Further assuming that given a tree, the priors on its terminal node parameters are independent. Therefore,

$$\begin{aligned} p(\mathbf{T}, \mathbf{M}, \tilde{\mathbf{v}}, \tau) &= \left[\prod_{j=1}^{n_T} p(T_j, M_j) \right] p(\tilde{\mathbf{v}} | \tau) p(\tau) \\ &= \left[\prod_{j=1}^{n_T} p(T_j) p(M_j | T_j) \right] p(\tilde{\mathbf{v}} | \tau) p(\tau) \\ &= \left[\prod_{j=1}^{n_T} \left[p(T_j) \prod_{k=1}^{m_j} p(\mu_{jk} | T_j) \right] \right] p(\tilde{\mathbf{v}} | \tau) p(\tau), \end{aligned}$$

where $M_j = (\mu_{j1}, \dots, \mu_{jm_j})$, m_j is the number of terminal nodes of tree T_j , and n_T is the number of trees in the sum.

For the prior $p(T_j)$, we follow a *tree generating stochastic process* of Chipman et al. (2012).

$$P(\text{split at depth } d) = \begin{cases} 1 & d = 0 \\ \frac{\alpha_T}{(1+d)^{\beta_T}}, & d > 0, \end{cases}$$

where $\alpha_T \in (0, 1)$, $\beta_T \in [0, \infty)$. In addition, we use uniform distribution for both of the distribution over the splitting variable and the distribution over the splitting rule.

Given a tree T_j , the prior on the terminal node parameters is a Gaussian distribution, $p(\mu_{jk}|T_j) \sim N(\mu_\mu, \sigma_\mu^2)$, for $k = 1, \dots, m_j$. Hyper-parameters μ_μ and σ_μ^2 are selected so that the overall effect induced by the prior distributions is in the interval $(y_{\min}, y_{\max})$ with high probability. As in Kindo et al. (2016), we use the transformation $\tilde{y} = \frac{y - y_{\min}}{y_{\max} - y_{\min}} - 0.5$, ensuring that the transformed response lies in the $(-0.5, 0.5)$ interval. Consequently, $p(\mu_{jk}|T_j) \sim N(0, \frac{1}{4\kappa^2 n_T})$, for $k = 1, \dots, m_j$. We set $\kappa = 2$, which has been found to yield good results, though it can be optimized through cross-validation.

Following standard Bayesian approach to CQR, we use the expression of ALD as a mixture of an exponential and a scaled normal distribution and obtain the following hierarchical models for BART-CQR:

$$\begin{aligned} y_i &= b_{\theta_k} + \sum_j g(x_i; T_j, M_j) + \xi_{1k} \tilde{v}_i \\ &\quad + \tau^{-1/2} \xi_{2k} \sqrt{\tilde{v}_i} z_i, \text{ for all } i \text{ such that } C_{ik} = 1, \\ \tilde{v}|\tau &\sim \prod_{i=1}^n \tau \exp(-\tau \tilde{v}_i), \\ \tau &\sim \tau^{a_\tau - 1} \exp(-b_\tau \tau), \\ \mathbf{z} &\sim \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z_i^2\right), \\ \omega &\sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K), \end{aligned} \quad (6)$$

where $\xi_{1k} = \frac{1-2\theta_k}{\theta_k(1-\theta_k)}$, and $\xi_{2k}^2 = \frac{2}{\theta_k(1-\theta_k)}$.

3 Posterior Inference

The posterior computation is carried out using a Gibbs sampling algorithm. Based on the hierarchical model in (6), the posterior updating scheme proceeds through the following six steps:

$$\begin{aligned} &f(\tilde{v}_i|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{v}_{(-i)}, \mathbf{b}, \mathbf{C}, \omega) \text{ for } i = 1, \dots, n \\ &f((T_j, M_j)|\mathbf{T}_{(-j)}, \mathbf{M}_{(-j)}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{v}, \mathbf{b}, \mathbf{C}, \omega) \\ &\text{for } j = 1, \dots, n_T \end{aligned}$$

$$\begin{aligned} &f(\tau|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tilde{v}, \mathbf{b}, \mathbf{C}, \omega) \\ &f(\mathbf{C}_i|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{v}, \mathbf{b}, \mathbf{C}_{-i}, \omega) \text{ for } i = 1, \dots, n \\ &f(\omega|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{v}, \mathbf{b}, \mathbf{C}) \\ &f(b_{\theta_k}|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{v}, b_{-\theta_k}, \mathbf{C}, \omega) \text{ for } k = 1, \dots, K. \end{aligned}$$

The full conditional distribution of $\tilde{v}_i$ follows a Generalized Inverse Gaussian distribution. The distribution of the precision parameter τ is Gamma, and the latent class assignments $\mathbf{C}_i$ are drawn from a Multinomial distribution. For the mixture weights ω , we use a Dirichlet distribution, and the intercept b_{θ_k} has a normal full conditional distribution.

The most challenging part is drawing the regression trees, (T_j, M_j) . Each pair is updated using a Bayesian backfitting strategy based on residuals from the other trees, similar to standard BART procedures. First define

$$\begin{aligned} r_{ij} &:= y_i - b_{\theta_k} - \sum_{l \neq j} g(\mathbf{x}_i; T_l, M_l) - \xi_{1k} \tilde{v}_i = g(\mathbf{x}_i; T_j, M_j) \\ &\quad + \tau^{-1/2} \xi_{2k} \sqrt{\tilde{v}_i} z_i, \\ &\text{for all } i \text{ such that } C_{ik} = 1. \end{aligned} \quad (7)$$

Then, draw from (T_j, M_j) is equivalent to draw from a single regression tree $r_{ij} = g(\mathbf{x}_i; T_j, M_j) + \tau^{-1/2} \xi_{2k} \sqrt{\tilde{v}_i} z_i$ for $i = 1, \dots, n$. In the tree T_j , assume that there are m_j terminal nodes and that n_l observations fall into terminal node l for $l = 1, \dots, m_j$ ($n_1 + \dots + n_{m_j} = n$). Let's consider the set of observations fall into terminal node l . Since for each y_i , there is corresponding C_{ik} such that $C_{ik} = 1$, we can consider a pair $\{(i, k) | k \text{ such that } C_{ik} = 1\}$ for each i . Denote the set of (i, k) 's that fall into terminal node l as G_l . Further, define $\mathbf{r}_j = (\mathbf{r}_{j,1}, \dots, \mathbf{r}_{j,m_j})$ with $\mathbf{r}_{j,l} = (r_{1,j,l}, \dots, r_{n_l,j,l})$, where $r_{i,j,l}$ be a observation that falls into l th terminal node in tree T_j for $i \in G_l$, and we have $r_{i,j,l} \sim N(\mu_{jl}, \tau^{-1} \xi_{2k}^2 \tilde{v}_i)$ for $(i, k) \in G_l$. Then the likelihood of the single tree in (7) is

$$\begin{aligned} &f(\mathbf{r}_j|\mathbf{X}, \mathbf{T}_j, \mathbf{M}_j, \tau, \tilde{v}, \mathbf{b}, \mathbf{C}, \omega) \\ &= \prod_{l=1}^{m_j} f(\mathbf{r}_{j,l}|\mathbf{X}_l, \mathbf{T}_j, \mathbf{M}_j, \tau, \tilde{v}_l, \mathbf{b}, \mathbf{C}, \omega), \end{aligned}$$

where

$$\begin{aligned} &f(\mathbf{r}_{j,l}|\mathbf{X}_l, \mathbf{T}_j, \mathbf{M}_j, \tau, \tilde{v}_l, \mathbf{b}, \mathbf{C}, \omega) \\ &= \prod_{(i,k) \in G_l} \frac{1}{\sqrt{2\pi \tau^{-1} \xi_{2k}^2 \tilde{v}_i}} \exp\left[-\frac{1}{2} \frac{(r_{i,j,l} - \mu_{jl})^2}{\tau^{-1} \xi_{2k}^2 \tilde{v}_i}\right]. \end{aligned}$$

Now, the draw from $f((T_j, M_j)|\mathbf{T}_{(-j)}, \mathbf{M}_{(-j)}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{v}, \mathbf{b}, \mathbf{C}, \omega)$ can be done by two successive steps as

$$f(T_j | \mathbf{r}_j, \tau, \tilde{v}, \mathbf{b}, \mathbf{C}, \omega), \quad (8)$$

$$f(M_j | T_j, \mathbf{r}_j, \tau, \tilde{v}, \mathbf{b}, \mathbf{C}, \omega). \quad (9)$$

For (8), we use Metropolis-Hastings algorithm. The formula of $f(T_j|\mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega)$ can be derived as

$$\begin{aligned} f(T_j|\mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) &\propto p(T_j) \\ &\int p(\mathbf{r}_j|M_j, T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) p(M_j|T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) dM_j \\ &= p(T_j) \prod_l \left\{ \left(\frac{1}{\sqrt{2\pi\tau^{-1}}} \right)^{n_l} \left(\prod_{(i,k) \in G_l} \xi_{2k}^{-1} \tilde{v}_i^{-1/2} \right) \right. \\ &\quad \exp \left(-\frac{1}{2} \sum_{(i,k) \in G_l} \frac{r_{i,j,l}^2}{\tau^{-1} \xi_{2k}^2 \tilde{v}_i} \right) \\ &\quad \times \sqrt{\frac{\tau^{-1}}{\sigma_\mu^2 \sum_{(i,k) \in G_l} \tilde{v}_i^{-1} \xi_{2k}^{-2} + \tau^{-1}}} \\ &\quad \left. \exp \left[\frac{\sigma_\mu^2 \left(\sum_{(i,k) \in G_l} r_{i,j,l} \tilde{v}_i^{-1} \xi_{2k}^{-2} \right)^2}{2\tau^{-1} \left(\tau^{-1} + \sigma_\mu^2 \sum_{(i,k) \in G_l} \tilde{v}_i^{-1} \xi_{2k}^{-2} \right)} \right] \right\}. \end{aligned} \quad (10)$$

Given an update tree, μ_{jl} , l th terminal node parameter at tree T_j in (9) can be drawn as:

$$\begin{aligned} f(\mu_{jl}|T_j, \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) &\propto \\ &\exp \left[- \left(\sum_{(i,k) \in G_l} \frac{1}{2\tau^{-1} \xi_{2k}^2 \tilde{v}_i} + \frac{1}{2\sigma_\mu^2} \right) \mu_{jl}^2 \right. \\ &\quad \left. + 2 \sum_{(i,k) \in G_l} \frac{r_{i,j,l}}{2\tau^{-1} \xi_{2k}^2 \tilde{v}_i} \mu_{jl} \right], \end{aligned}$$

which is a Gaussian distribution.

Complete derivations and the step-by-step sampling scheme are provided in Appendix A.

After running the algorithm sufficiently long after the burn-in period, we obtain a sequence of posterior draws of f^* ;

$$f^*(x) = \sum_j g(x; T_j^*, M_j^*),$$

denoted as $f_1^*, \dots, f_K^*$. The final estimate of $f(x)$ at a given x is then taken as the average of $f_1^*, \dots, f_K^*$, as in Chipman et al. (2012).

4 Simulations

We consider Friedman's five dimensional test function (Friedman 1991) to illustrate various features of the proposed method on simulated data. We construct data by simulating values of $\mathbf{x} = (x_1, \dots, x_p)$, where $x_1, \dots, x_p \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}$

(0, 1), and

$$\begin{aligned} y = f(\mathbf{x}) + \epsilon &= 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 \\ &\quad + 10x_4 + 5x_5 + \epsilon. \end{aligned}$$

We consider various error distributions to demonstrate the superiority of the proposed method:

- Normal distribution: $\epsilon \sim N(0, 1)$.
- Contaminated Normal distribution: $\epsilon = \epsilon_1 + \zeta_1 \epsilon_2 + \zeta_2 \epsilon_3$, where $\zeta_j \sim \text{Bern}(0.15)$, $j = 1, 2$, being independent of $Y, \epsilon_1, \epsilon_2$, and ϵ_3 , $\epsilon_1 \sim N(0, 1)$, $\epsilon_2 \sim N(0, 9^2)$, and $\epsilon_3 \sim \text{ALD}(\tau = 1, \theta = 0.9)$.
- t -distribution: $\epsilon \sim t(2)$.
- Asymmetric Laplace distribution: $\epsilon \sim \text{ALD}(\tau = 1, \theta)$ with $\theta = 0.1, 0.9$.

The model evaluation metric used is the root mean squared error (RMSE), given by

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(\mathbf{x}_i) - \hat{y}_i)^2},$$

where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ is the i th covariate, $f(\mathbf{x}_i)$ is the true regression function used to generate the data and $\hat{y}_i$ is the model's prediction.

We compare the proposed composite quantile BART model (BART-CQR) with quantile BART (QBART) at quantile levels 25%, 50%, and 75% (Kindo et al. 2016), BART (Chipman et al. 2012), and composite quantile regression (CQ regression) (Huang and Chen 2015). For tree methods, we set the number of trees $n_T = 200$, and for the composite quantile methods, we use $K = 9$ quantile levels, $\theta_k = \frac{k}{K+1}$ for $k = 1, \dots, K$. For all methods, we set 3000 burn-in steps and use 8000 iterations.

We set $n = 200$ and $p = 30$ in each simulation and use 100 observations as the training set, with the remaining observations designated for the test set. We run 200 independent replications, and the test dataset performance is presented in Table 1.

The simulation results demonstrate that the BART-CQR model consistently outperforms other methods, particularly under heavy-tailed or skewed error distributions. As expected, classical BART performs best under normally distributed errors, where it is well suited; however, its performance deteriorates significantly under non-Gaussian settings. Notably, BART-CQR performs comparably well under normal errors, indicating its effectiveness even in well-behaved scenarios. Under the contaminated normal distribution and $t(2)$ distribution, BART-CQR also yields the best test RMSE, highlighting its robustness. Finally, with

Table 1 RMSE (standard error in parentheses) over 200 replications for Friedman’s five dimensional test function data with different error distributions. Bold face indicates the best performance

	BART-CQR	QBART (25%)	QBART (50%)	QBART (75%)	BART	CQ regression
Normal	1.800 (0.187)	2.641 (0.347)	2.736 (0.324)	2.644 (0.289)	1.739 (0.175)	2.591 (0.240)
Contaminated Normal	2.712(0.427)	8.285(1.630)	8.411(1.660)	7.276(1.487)	3.729(0.629)	3.245(0.440)
$t(2)$	1.970(0.206)	3.826(1.541)	4.076(1.320)	3.624(0.826)	2.325(0.811)	2.724(0.251)
ALD ($\theta = 0.1$)	9.317(1.347)	14.228(2.105)	16.33(2.427)	16.774(2.501)	10.098(1.223)	9.689(1.181)
ALD ($\theta = 0.9$)	8.886(1.299)	16.343(2.534)	15.976(2.67)	13.922(2.196)	9.862(1.236)	9.372(1.014)

the ALD at $\theta = 0.1$ and $\theta = 0.9$, BART-CQR maintains the lowest test RMSE, confirming its ability to handle complex and asymmetric error structures effectively. Interestingly, CQ regression outperforms QBART even though the underlying regression function is nonlinear. This may be attributed to the composite nature of the composite quantile loss, which aggregates information across multiple quantiles, thereby stabilizing estimation and reducing variance. In contrast, QBART focuses on individual quantiles, which may be more sensitive to noise and outliers, particularly under heavy-tailed or asymmetric error distributions.

To examine the sensitivity of BART-CQR to the prior specification on the error precision parameter τ , we conducted a prior sensitivity analysis using three Gamma prior settings. The results, presented in Appendix B, indicate that the performance of BART-CQR remains robust across a wide range of prior choices.

We conduct additional simulations under the contaminated normal setting with increased training sample sizes to further evaluate the scalability of BART-CQR. The results, reported in Appendix C, show that BART-CQR’s performance improves as the sample size increases, consistently achieving the lowest RMSE compared to competitors.

We further investigate the effect of heteroscedastic errors and correlated input variables under the following model with $n = 200$ and $p = 30$:

$$y = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + \sigma(\mathbf{x})\epsilon,$$

where $\mathbf{x} := (x_1, \dots, x_p)^T \sim N_p(\mathbf{0}, \Sigma)$, $\Sigma = (\sigma_{ij})_{p \times p}$ with $\sigma_{ij} = 0.5^{|i-j|}$, and $\sigma(\mathbf{x}) = \{(1 + 2\mathbf{x}^T)/3\}\mathbf{1}$, and $\epsilon \sim N(0, 1)$. For each of the 200 replications, we randomly split the data into a training set of size 100 and a test set of size 100. The predictive performance was evaluated using the test set only. The results in Table 2 show that BART-CQR consistently achieves the lowest RMSE, demonstrating superior predictive performance under both heteroscedasticity and correlated covariates.

5 Real Data Examples

For the real data analysis, we consider three benchmark/public datasets:

Ozone Data: This dataset records ozone levels (in parts per billion) in New York from May to September 1973. The predictors include solar radiation level, wind speed, maximum daily temperature, month, and day of measurement. After removing observations with missing values, we have $n = 153$ observations.

Auto Insurance Data: This dataset consists of $n = 2812$ auto insurance policyholders with 56 predictors along with an aggregate paid claim amount. Examples of predictors include the driver’s age, driver’s income, vehicle use (commercial or non-commercial), vehicle type (one of six categories), and the driver’s gender. The response variable is the aggregate claim amount, which is skewed with a significant number of policyholders having zero claims. For non-zero claims, the reported amounts tend to be larger.

Boston Housing Data: This dataset includes $n = 506$ samples. We examine the relationship between the log-transformed corrected median value of owner-occupied housing (in \$1000), denoted as medv , and 13 explanatory variables: crim (per capita crime rate by town), zn (proportion of residential land zoned for lots over 25,000 sq.ft), indus (proportion of non-retail business acres per town), chas (Charles River dummy variable), nox (nitrogen oxides concentration), rm (average number of rooms per dwelling), age (proportion of owner-occupied units built prior to 1940), dis (weighted mean of distances to five Boston employment centers), rad (index of accessibility to radial highways), tax (full-value property-tax rate per \$10,000), ptratio (pupil-teacher ratio by town), lstat (percentage of lower status of the population).

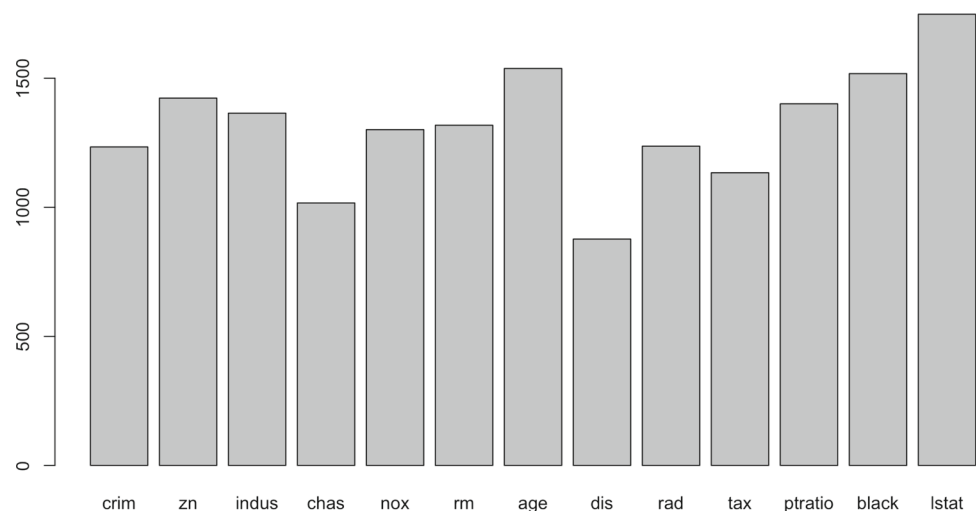
These three datasets are available in the R packages *datasets* (R Core Team 2024), *HDTweedie* (Qian et al. 2022), and *MASS* (Venables and Ripley 2002), respectively. To evaluate predictive performance, we apply 5-fold cross-validation to each dataset. Specifically, the data are randomly partitioned into five nearly equal-sized folds. In each iteration, four of the five folds are used for training, and the remaining fold is used for testing. Predictive RMSE is com-

Table 2 RMSE (standard error in parentheses) over 200 replications for heteroscedastic error model. Bold face indicates the best performance

BART-CQR	QBART (25%)	QBART (50%)	QBART (75%)	BART	CQ regression
16.623(3.676)	21.347(3.629)	24.568(4.399)	23.345(5.398)	17.419(4.282)	31.297(4.623)

Table 3 Real data: Test data average RMSE based on 5 folds cross-validation

	BART-CQR	QBART (25%)	QBART (50%)	QBART (75%)	CQ regression
Ozone Data	16.920	20.587	23.769	21.327	26.149
Auto Insurance Data	8.457	9.352	8.682	9.819	8.584
Boston Housing Data	0.156	0.192	0.190	0.189	0.186

Fig. 1 Variable used count in the BART-CQR for Boston housing data

puted on each test fold, and the average RMSE across all five folds is reported as the final performance measure. As in the simulation study, we set the number of trees to $n_T = 200$, and for the composite quantile methods, we use $K = 9$ quantile levels.

Table 3 summarizes the results. For all data, the BART-CQR provides the smallest RMSE, implying that the proposed method works well with complex real data sets. While the composite quantile regression model may work well if the variables are linearly related, it may collapse under complicated structures. On the other hand, BART models perform well with various structured data, but we need to determine the proper quantile level for the QBART. The proposed composite quantile BART strikes a balance and provides the best performance in these datasets.

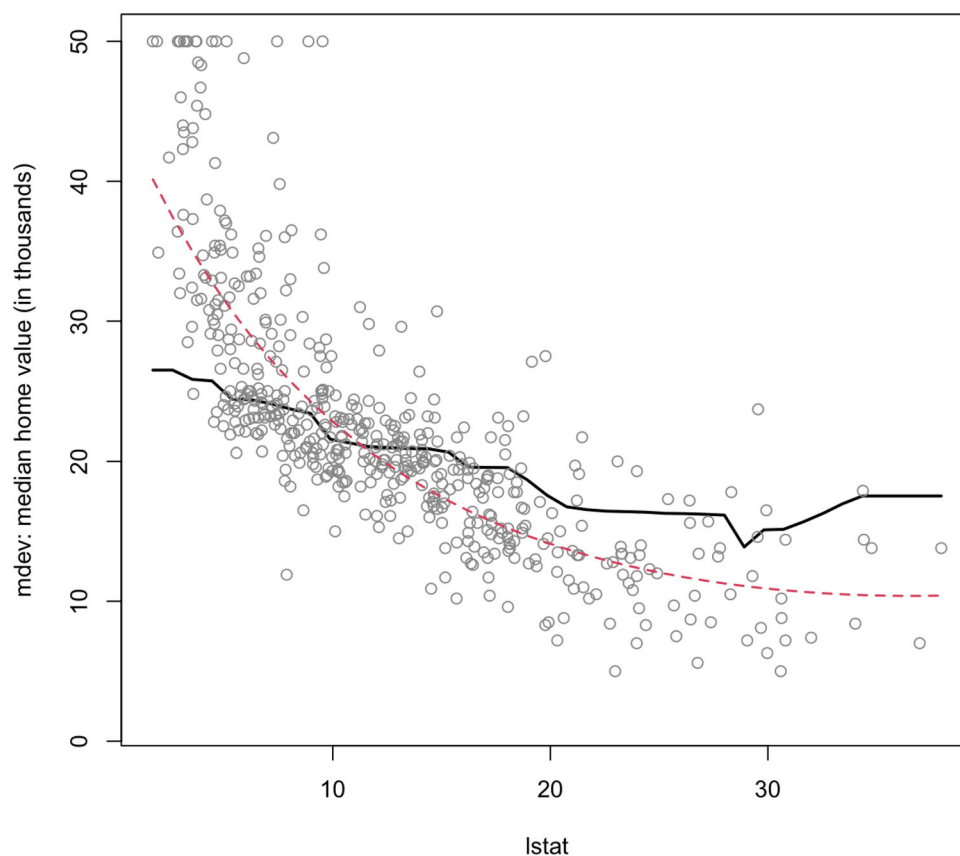
For the interpretation, we consider the effect of predictors on the outcome. However, tree models do not directly provide a summary of the effect of a single predictor, or a subset of predictors, on the outcome. We first examine how many times each variable appeared in the collection of trees, which provides a summary similar to the variable importance plot used in boosting and random forests. For simplicity and conciseness of the paper, we report results only for the Boston

housing data, among the three real datasets we considered. Figure 1 shows the barplot of the counts in BART-CQR for Boston housing data. We observe that *lstat* appears most frequently in the trees, highlighting the significant impact of socio-economic status on housing prices.

Furthermore, as suggested by Chipman et al. (2014), we use Friedman's partial dependence function (Friedman 2001) to summarize the marginal effect due to a subset of the predictors, $\mathbf{x}_S$, by aggregating over the predictors in the complement set, $\mathbf{x}_C$, i.e., $\mathbf{x} = [\mathbf{x}_S, \mathbf{x}_C]$. The marginal dependence function is defined by fixing $\mathbf{x}_S$ while aggregating over the observed settings of the complement predictors in the data set: $f(\mathbf{x}_S) = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_S, \mathbf{x}_{iC})$.

Figure 2 summarizes the marginal effect of *lstat* on *mdev* while aggregating over the other predictors with Friedman's partial dependence function. We observe a negative effect on *mdev*, shown by the black solid line, which implies that less affluent neighborhoods have lower home values. Compared to the quadratic regression results, shown by the red dashed line, BART-CQR provides a more robust fitted line and also well-captures the complex non-linear relationship between the predictors and the outcome.

Fig. 2 The Boston housing data: the marginal effect of `lstat` on `medv` while aggregating over the other covariates with Friedman's partial dependence function. The marginal estimate from BART-CQR is shown by the black solid line, and the red dashed line comes from the linear regression model where a quadratic effect of `lstat` with respect to the logarithm of `medv` is assumed



6 Conclusion and Discussion

In this paper, we proposed a novel Bayesian framework, BART-CQR, which integrates BART with the CQR approach. This method is designed to handle complex, nonlinear relationships and heavy-tailed error distributions, extending the flexibility and robustness of existing quantile-based models. We developed a fully Bayesian hierarchical formulation for BART-CQR and derived an efficient Gibbs sampler for posterior inference. Through comprehensive simulation studies and real data applications, we demonstrated that BART-CQR consistently outperforms classical BART, quantile BART and standard linear composite quantile regression in terms of predictive accuracy. Compared to quantile BART, BART-CQR eliminates the challenge of quantile level selection. Additionally, it offers greater modeling flexibility than standard CQR by capturing nonlinear effects and interactions via tree ensembles. Thus, the proposed BART-CQR method inherits the strengths of both the BART framework and the composite quantile approach, offering a powerful alternative for analyzing complex, non-Gaussian data in moderately sized regression problems with up to approximately 3000 observations.

The proposed approach has a couple of limitations. Although there is room to improve computational efficiency

through careful optimization, BART-CQR requires more time than classical BART and quantile BART due to the complexity of the tree ensemble model and MCMC sampling, which may limit its scalability to very large datasets. Moreover, the current method assumes independent and identically distributed errors; its performance under dependent or longitudinal data settings remains to be investigated. In addition, this work does not explore model selection or variable selection properties of BART-CQR, which could be important in high-dimensional settings. Investigating these aspects both theoretically and empirically constitutes an important direction for future research.

Several promising directions exist for future research. Extensions could include incorporating recent advances in BART, such as softBART (Linero and Yang 2018) and BART models with random effects for hierarchical data (Tan and Roy 2019). Furthermore, model selection criteria, theoretical properties of the BART-CQR estimator, and applications to specific domains such as finance, epidemiology, and environmental science warrant further investigation.

Appendix A Posterior updating scheme

The complete likelihood is

$$\begin{aligned} f(\mathbf{y}|\mathbf{X}, \mathbf{T}, \mathbf{M}, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) \\ = \prod_{i=1}^n \prod_{k=1}^K \left(\frac{1}{\sqrt{2\pi\tau^{-1}\xi_{2k}^2\tilde{v}_i}} \right)^{C_{ik}} \\ \exp \left[-\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^K \frac{C_{ik}(y_i - b_{\theta_k} - \sum_j g(x_i; T_j, M_j) - \xi_{1k}\tilde{v}_i)^2}{\tau^{-1}\xi_{2k}^2\tilde{v}_i} \right], \end{aligned}$$

and the posterior updating scheme cycles can be done through the following six posterior draws:

$$f(\tilde{v}_i|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{v}_{(-i)}, \mathbf{b}, \mathbf{C}, \omega) \quad (\text{A1})$$

$$f((T_j, M_j)|\mathbf{T}_{(-j)}, \mathbf{M}_{(-j)}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) \quad (\text{A2})$$

$$f(\tau|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) \quad (\text{A3})$$

$$f(\mathbf{C}_i|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}_{-i}, \omega) \quad (\text{A4})$$

$$f(\omega|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}) \quad (\text{A5})$$

$$f(b_{\theta_k}|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, b_{-\theta_k}, \mathbf{C}, \omega). \quad (\text{A6})$$

For (A1), we have $f(\tilde{v}_i|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{v}_{(-i)}, \mathbf{b}, \mathbf{C}, \omega) \propto v_i^{-1/2} \exp \left\{ -\frac{1}{2} \left(\delta_{1i} \tilde{v}_i^{-1} + \delta_{2i} \tilde{v}_i \right) \right\}$, where $\delta_{1i} := \sum_k \frac{C_{ik}(y_i - b_{\theta_k} - \sum_j g(x_i; T_j, M_j))^2}{\tau^{-1}\xi_{2k}^2}$ and $\delta_{2i} = \sum_k \frac{C_{ik}\xi_{1k}^2}{\xi_{2k}^2\tau^{-1}} + 2\tau$. Therefore, we sequentially samples $\tilde{v}_i$ from a Generalized Inverse Gaussian distribution.

For (A3), we derive that

$$f(\tau|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) \propto \tau^{n/2+n+\frac{\alpha}{2}-1} \exp \left\{ -\tau \left[\sum_{i=1}^n \sum_{k=1}^K \frac{C_{ik}(y_i - b_{\theta_k} - \sum_j g(x_i; T_j, M_j) - \xi_{1k}\tilde{v}_i)^2}{2\xi_{2k}^2\tilde{v}_i} + \sum_{i=1}^n \tilde{v}_i + b_\tau \right] \right\},$$

which is a Gamma distribution, and for $f(\mathbf{C}_i|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}_{-i}, \omega)$ in (A4),

$$\begin{aligned} f(\mathbf{C}_i|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}_{-i}, \omega) \\ \propto \prod_{k=1}^K \left\{ \left(\frac{1}{\sqrt{2\pi\tau^{-1}\xi_{2k}^2\tilde{v}_i}} \right)^{C_{ik}} \right. \\ \left. \exp \left[-\frac{1}{2} \frac{C_{ik}(y_i - b_{\theta_k} - \sum_j g(x_i; T_j, M_j) - \xi_{1k}\tilde{v}_i)^2}{\tau^{-1}\xi_{2k}^2\tilde{v}_i} \right] \right\} \\ \prod_{k=1}^K \omega_k^{C_{ik}} \propto \prod_{k=1}^K \left\{ \left(\frac{\omega_k}{\xi_{2k}} \right) \right. \\ \left. \exp \left[-\frac{1}{2} \frac{(y_i - b_{\theta_k} - \sum_j g(x_i; T_j, M_j) - \xi_{1k}\tilde{v}_i)^2}{\tau^{-1}\xi_{2k}^2\tilde{v}_i} \right] \right\}^{C_{ik}}, \end{aligned}$$

which is a Multinomial($1, \hat{p}_1, \dots, \hat{p}_K$), where

$$\hat{p}_k = \frac{\left(\frac{\omega_k}{\xi_{2k}} \right) \exp \left[-\frac{1}{2} \frac{(y_i - b_{\theta_k} - \sum_j g(x_i; T_j, M_j) - \xi_{1k}\tilde{v}_i)^2}{\tau^{-1}\xi_{2k}^2\tilde{v}_i} \right]}{\sum_k \left\{ \left(\frac{\omega_k}{\xi_{2k}} \right) \exp \left[-\frac{1}{2} \frac{(y_i - b_{\theta_k} - \sum_j g(x_i; T_j, M_j) - \xi_{1k}\tilde{v}_i)^2}{\tau^{-1}\xi_{2k}^2\tilde{v}_i} \right] \right\}}.$$

The draw in (A5) is Dirichlet($n_1 + \alpha_1, \dots, n_K + \alpha_K$)

$$f(\omega|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}) \propto \prod_{k=1}^K \omega_k^{\alpha_k + n_k}, \text{ where } n_k := \sum_i C_{ik}.$$

For the intercept (A6),

$$\begin{aligned} f(b_{\theta_k}|\mathbf{T}, \mathbf{M}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, b_{-\theta_k}, \mathbf{C}, \omega) \propto \\ \exp \left(-\frac{1}{2} \sum_i \frac{C_{ik}}{\tau^{-1}\xi_{2k}^2\tilde{v}_i} b_{\theta_k}^2 + \sum_i \frac{C_{ik}\tilde{r}_{ik}}{\tau^{-1}\xi_{2k}^2\tilde{v}_i} b_{\theta_k} \right), \end{aligned}$$

where $\tilde{r}_{ik} := y_i - \sum_j g(x_i; T_j, M_j) - \xi_{1k}\tilde{v}_i$, and it is a normal distribution.

Now, for the draw (A2), first define

$$\begin{aligned} r_{ij} := y_i - b_{\theta_k} - \sum_{l \neq j} g(\mathbf{x}_i; T_l, M_l) - \xi_{1k}\tilde{v}_i = g(\mathbf{x}_i; T_j, M_j) \\ + \tau^{-1/2}\xi_{2k}\sqrt{\tilde{v}_i}z_i, \\ \text{for } y_i \in \text{cluster } k. \end{aligned} \quad (\text{A7})$$

Then, draw from (A2) is equivalent to draw from a single regression tree $r_{ij} = g(\mathbf{x}_i; T_j, M_j) + \tau^{-1/2}\xi_{2k}\sqrt{\tilde{v}_i}z_i$ for

$i = 1, \dots, n$. In the tree T_j , assume that there are m_j terminal nodes and n_l observations fall into terminal node l for $l = 1, \dots, m_j$ ($n_1 + \dots + n_{m_j} = n$). Let's consider the set of observations fall into terminal node l . Since for each y_i , there is corresponding C_{ik} such that $C_{ik} = 1$, we can consider a pair $\{(i, k) | k \text{ such that } C_{ik} = 1\}$ for each i . Denote the set of (i, k) 's that fall into terminal node l as G_l . Further, define $\mathbf{r}_j = (\mathbf{r}_{j,1}, \dots, \mathbf{r}_{j,m_j})$ with $\mathbf{r}_{j,l} = (r_{1,j,l}, \dots, r_{n_l,j,l})$, where $r_{i,j,l}$ be a observation that falls into l th terminal node in tree T_j for $i \in G_l$, and we have $r_{i,j,l} \sim N(\mu_{jl}, \tau^{-1}\xi_{2k}^2\tilde{v}_i)$ for $(i, k) \in G_l$. Then the likelihood of the single tree in (A7) is

$$f(\mathbf{r}_j | \mathbf{X}, \mathbf{T}_j, \mathbf{M}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) \\ = \prod_{l=1}^{m_j} f(\mathbf{r}_{j,l} | \mathbf{X}_l, \mathbf{T}_j, \mathbf{M}_j, \tau, \tilde{\mathbf{v}}_l, \mathbf{b}, \mathbf{C}, \omega),$$

where

$$f(\mathbf{r}_{j,l} | \mathbf{X}_l, \mathbf{T}_j, \mathbf{M}_j, \tau, \tilde{\mathbf{v}}_l, \mathbf{b}, \mathbf{C}, \omega) \\ = f(\mathbf{r}_{j,l} | \mu_{jl}, \tau, \tilde{\mathbf{v}}_l, \mathbf{b}, \mathbf{C}, \omega) \\ = \prod_{(i,k) \in G_l} \frac{1}{\sqrt{2\pi\tau^{-1}\xi_{2k}^2\tilde{v}_i}} \exp\left[-\frac{1}{2} \frac{(r_{i,j,l} - \mu_{jl})^2}{\tau^{-1}\xi_{2k}^2\tilde{v}_i}\right].$$

Now, the draw from $f((T_j, M_j) | \mathbf{T}_{(-j)}, \mathbf{M}_{(-j)}, \mathbf{Y}, \mathbf{X}, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega)$ can be done by two successive steps as

$$f(T_j | \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega), \quad (\text{A8})$$

$$f(M_j | T_j, \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega). \quad (\text{A9})$$

For (A8), we use Metropolis-Hastings algorithm. We first derive a formula of $f(T_j | \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega)$:

$$f(T_j | \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) \propto p(T_j) \\ \int p(\mathbf{r}_j | M_j, T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) p(M_j | T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) dM_j \\ = p(T_j) \prod_l \left[\int \left(\prod_{(i,k) \in G_l} p(r_{i,j,l} | M_j, T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) \right) \right.$$

$$\times p(\mu_{jl} | T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) d\mu_{jl} \Big] \\ = p(T_j) \prod_l \left[\int \left\{ \prod_{(i,k) \in G_l} \frac{1}{\sqrt{2\pi\tau^{-1}\xi_{2k}^2\tilde{v}_i}} \right. \right. \\ \exp\left[-\frac{1}{2} \frac{(r_{i,j,l} - \mu_{jl})^2}{\tau^{-1}\xi_{2k}^2\tilde{v}_i}\right] \Big\} \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \\ \exp\left(-\frac{\mu_{jl}^2}{2\sigma_\mu^2}\right) d\mu_{jl} \Big] \\ = p(T_j) \prod_l \left\{ \left(\frac{1}{\sqrt{2\pi\tau^{-1}}} \right)^{n_l} \left(\prod_{(i,k) \in G_l} \xi_{2k}^{-1} \tilde{v}_i^{-1/2} \right) \right. \\ \exp\left(-\frac{1}{2} \sum_{(i,k) \in G_l} \frac{r_{i,j,l}^2}{\tau^{-1}\xi_{2k}^2\tilde{v}_i}\right) \\ \times \sqrt{\frac{\tau^{-1}}{\sigma_\mu^2 \sum_{(i,k) \in G_l} \tilde{v}_i^{-1} \xi_{2k}^{-2} + \tau^{-1}}} \\ \left. \exp\left[\frac{\sigma_\mu^2 \left(\sum_{(i,k) \in G_l} r_{i,j,l} \tilde{v}_i^{-1} \xi_{2k}^{-2} \right)^2}{2\tau^{-1} \left(\tau^{-1} + \sigma_\mu^2 \sum_{(i,k) \in G_l} \tilde{v}_i^{-1} \xi_{2k}^{-2} \right)} \right] \right\}. \quad (\text{A10})$$

Then, to draw from $f(T_j | \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega)$, we first obtain a new candidate tree T_j^* with a acceptance rate,

$$\tilde{\alpha} = \min \left(1, \frac{f(T_j^* | \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) q(T_j | T_j^*)}{f(T_j | \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) q(T_j^* | T_j)} \right) \\ = \min \left(1, \frac{p(T_j^*) \int p(\mathbf{r}_j | M_j^*, T_j^*, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) p(M_j^* | T_j^*, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) dM_j q(T_j | T_j^*)}{p(T_j) \int p(\mathbf{r}_j | M_j, T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) p(M_j | T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) dM_j q(T_j^* | T_j)} \right) \\ = \min \left(1, \frac{p(\mathbf{r}_j | T_j^*, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) p(T_j^*) q(T_j | T_j^*)}{p(\mathbf{r}_j | T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) p(T_j) q(T_j^* | T_j)} \right).$$

The transition kernel $q(\cdot)$ assigns probabilities of 0.25, 0.25, 0.40, and 0.10 to the moves GROW, PRUNE, SWAP, and CHANGE respectively. Compared to quantile BART (Kindo et al. 2016), we use the same transition kernel and priors for the trees, with the only difference being the likelihood function.

For example, in the GROW case, the likelihood ratio can be computed as

$$\begin{aligned} \frac{p(\mathbf{r}_j|T_j^*, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega)}{p(\mathbf{r}_j|T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega)} &= \frac{\prod_{l \in \text{two children}} \left[\int \prod_{i \in G_l} p(r_{i,j,l}|M_j, T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) p(\mu_{jl}|T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) d\mu_{jl} \right]}{\prod_{l \in \text{parent}} \left[\int \prod_{i \in G_l} p(r_{i,j,l}|M_j, T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) p(\mu_{jl}|T_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) d\mu_{jl} \right]} \\ &= \frac{(\text{A10}) \text{ of left child node} \times (\text{A10}) \text{ of right child node}}{(\text{A10}) \text{ of parent node}} = \sqrt{\frac{\tau^{-1}(\tau^{-1} + \sigma_\mu^2 B_P)}{(\tau^{-1} + \sigma_\mu^2 B_L)(\tau^{-1} + \sigma_\mu^2 B_R)}} \\ &\quad \times \exp \left\{ \frac{\sigma_\mu^2}{2\tau^{-1}} \left(\frac{A_L^2}{\tau^{-1} + \sigma_\mu^2 B_L} + \frac{A_R^2}{\tau^{-1} + \sigma_\mu^2 B_R} - \frac{A_P^2}{\tau^{-1} + \sigma_\mu^2 B_P} \right) \right\}, \end{aligned}$$

where subscript P, L and R denote “parent”, “left”, and “right” nodes. $B_P := \sum_{(i,k) \in G_P} \tilde{v}_i^{-1} \xi_{2k}^{-2}$ and $A_P := \sum_{(i,k) \in G_P} r_{i,j,l} \tilde{v}_i^{-1} \xi_{2k}^{-2}$, where G_P is the set of (i, k) s fall into parent node. Similarly define A_L, B_L, A_R , and B_R .

Given an update tree, μ_{jl} , l th terminal node parameter at tree T_j in (A9) can be drawn as:

$$\begin{aligned} f(\mu_{jl}|T_j, \mathbf{r}_j, \tau, \tilde{\mathbf{v}}, \mathbf{b}, \mathbf{C}, \omega) &\propto \exp \left[- \sum_{(i,k) \in G_l} \frac{(r_{i,j,l} - \mu_{jl})^2}{2\tau^{-1} \xi_{2k}^2 \tilde{v}_i} \right] \\ &\quad \exp \left(- \frac{\mu_{jl}^2}{2\sigma_\mu^2} \right) \\ &\propto \exp \left[- \left(\sum_{(i,k) \in G_l} \frac{1}{2\tau^{-1} \xi_{2k}^2 \tilde{v}_i} + \frac{1}{2\sigma_\mu^2} \right) \mu_{jl}^2 \right. \\ &\quad \left. + 2 \sum_{(i,k) \in G_l} \frac{r_{i,j,l}}{2\tau^{-1} \xi_{2k}^2 \tilde{v}_i} \mu_{jl} \right], \end{aligned}$$

which is a Gaussian distribution.

Appendix B Prior Sensitivity Analysis for the Error Precision Parameter τ

As discussed in the main text, BART-CQR assumes an ALD for the error with precision parameter τ for which we place a Gamma prior:

$$\tau \sim \tau^{a_\tau-1} \exp(-b_\tau \tau).$$

To assess sensitivity to the hyperparameters (a_τ, b_τ) , we considered three prior configurations reflecting conservative, default, and aggressive prior beliefs (Figure 3):

- Conservative: $(a_\tau = 1, b_\tau = 2)$
- Default: $(a_\tau = 2, b_\tau = 1)$
- Aggressive: $(a_\tau = 10, b_\tau = 1)$

These choices span a wide range of assumptions regarding the concentration and spread of the ALD errors. Our empirical results suggest that the posterior estimates and predictive performance of BART-CQR remain stable across these prior settings, indicating that the method is not highly sensitive to the calibration of τ (see Table 4). Therefore, in contrast to standard BART, our model does not require fine-tuned calibration of the prior on τ , although it remains flexible for users who wish to encode informative prior beliefs.

Appendix C Additional Simulation: Effect of Increasing Sample Size under Contaminated Normal Errors

We conducted simulations under the contaminated normal setting using increased training sample sizes. Specifically, we used test sizes of 100, 100, and 200 for training sizes of 200, 400, and 800, respectively. As shown in Table 5, the performance of BART-CQR improves as the sample size increases, with notably lower RMSE at $n = 1000$ compared to $n = 200$. This trend confirms that the strength of our approach becomes more evident with larger sample sizes. Although performance improves across all methods as the sample size increases, BART-CQR consistently achieves the lowest RMSE across all settings, demonstrating its robustness and efficiency even in larger sample scenarios.

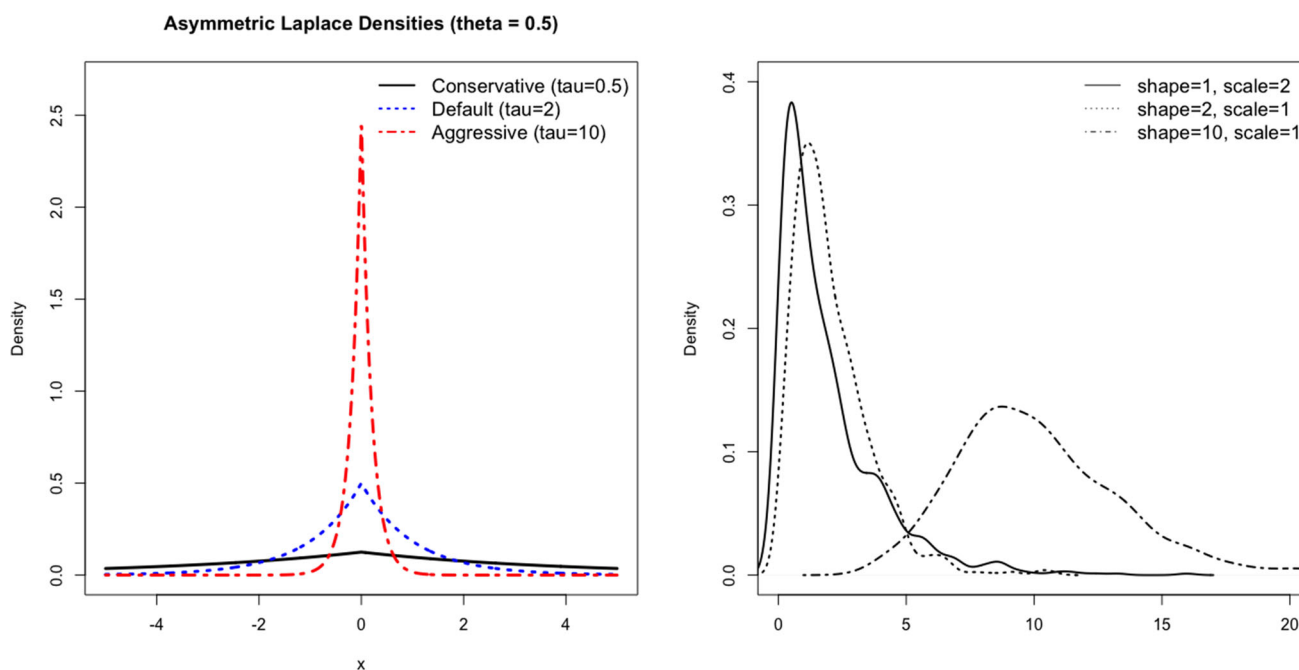


Fig. 3 (Left) Density functions of the ALD with $\theta = 0.5$ under three different values of the precision parameter τ . (Right) Prior densities of the precision parameter τ under three different Gamma priors: conser-

vative ($a_\tau = 1, b_\tau = 2$), default ($a_\tau = 2, b_\tau = 1$), and aggressive ($a_\tau = 10, b_\tau = 1$). These priors reflect increasing levels of certainty about the error precision.

Table 4 RMSE of BART-CQR under different Gamma priors on τ , based on simulations using Friedman's five dimensional test function with various error distributions

	$(a_\tau = 1, b_\tau = 2)$	$(a_\tau = 2, b_\tau = 1)$	$(a_\tau = 10, b_\tau = 1)$
Normal	1.809(0.174)	1.780(0.162)	1.776(0.165)
Contaminated Normal	2.196(0.208)	2.217(0.201)	2.197(0.194)
$t(2)$	1.947(0.196)	1.938(0.209)	1.915(0.213)
ALD ($\theta = 0.1$)	8.611(1.004)	8.385(0.987)	8.461(0.949)
ALD ($\theta = 0.9$)	8.169(1.065)	7.924(0.910)	7.922(0.94)

Table 5 RMSE (standard error in parentheses) over 100 replications under the contaminated normal setting for increasing training sample sizes (100, 400, and 800). Corresponding test sizes were 100, 100, and 200, respectively

	BART-CQR	QBART (25%)	QBART (50%)	QBART (75%)	BART	CQ regression
$n = 200$	2.712(0.427)	8.285(1.630)	8.411(1.660)	7.276(1.487)	3.729(0.629)	3.245(0.440)
$n = 500$	1.722(0.185)	8.580(1.172)	7.680(0.943)	6.309(0.755)	3.333(0.569)	3.050(0.310)
$n = 1000$	1.618(0.245)	7.549(0.692)	6.151(0.616)	4.868(0.413)	3.206(0.363)	2.864(0.274)

Acknowledgements Research of M. St. Ville, and Z. Chen was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), and Lim's research was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (2022R1F1A1074134).

Author Contributions Y.L: first author, methodology, computation and writing. R.L and M.V: methodology, conceptualization and writing. C. Z: corresponding author, methodology, conceptualization and writing.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

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 indi-

cate 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

- Alhamzawi, R.: Bayesian analysis of composite quantile regression. *Stat. Biosci.* **8**(2), 358–373 (2016)
- Basak, P., Linero, A., Sinha, D., Lipsitz, S.: Semiparametric analysis of clustered interval-censored survival data using soft Bayesian additive regression trees (SBART). *Biometrics* **78**(3), 880–893 (2022)
- Baumeister, C., Huber, F., Marcellino, M.: Risky Oil: It's All in the Tails. National Bureau of Economic Research (2024)
- Benoit, D.F., Van den Poel, D.: bayesQR: a bayesian approach to quantile regression. *J. Stat. Softw.* **76**, 1–32 (2017)
- Cao, T., Wu, J., Wang, Y.G.: An adaptive trimming approach to Bayesian additive regression trees. *Complex Intel. Syst.* **10**(5), 6805–6823 (2024)
- Chen, K., Müller, H.G.: Conditional quantile analysis when covariates are functions, with application to growth data. *J. R. Stat. Soc. Ser. B Stat Methodol.* **74**(1), 67–89 (2012)
- Chipman, H., George, E., Hahn, R., McCulloch, R., Pratola, M., Sparapani, R.: Bayesian additive regression trees, computational approaches, pp. 1–23. *Statistics Reference Online, Wiley StatsRef* (2014)
- Chipman, H.A., George, E.I., McCulloch, R.E.: BART: bayesian additive regression trees. *Ann. Appl. Stat.* **6**(1), 266–298 (2012)
- Clark, T.E., Huber, F., Koop, G., Marcellino, M., Pfarrhofer, M.: Tail forecasting with multivariate Bayesian additive regression trees. *Int. Econ. Rev.* **64**(3), 979–1022 (2023)
- Clark, T.E., Huber, F., Koop, G., Marcellino, M., Pfarrhofer, M.: Investigating growth-at-risk using a multicountry nonparametric quantile factor model. *J. Bus. Econ. Stat.* **42**(4), 1302–1317 (2024)
- Fitzenberger, B., Koenker, R., Machado, J.A.: *Economic applications of quantile regression*. Springer Science & Business Media (2013)
- Friedman, J.H.: Multivariate adaptive regression splines. *Ann. Stat.* **19**(1), 1–67 (1991)
- Friedman, J.H.: Greedy function approximation: a gradient boosting machine. *Annals of statistics*, p. 1189–1232 (2001)
- Galvao, A.F., Kato, K.: Smoothed quantile regression for panel data. *Journal of econometrics*. **193**(1), 92–112 (2016)
- Hao, L., Naiman, D.Q.: *Quantile regression*. No. 149, Sage; (2007)
- Haugen, M.A., Stein, M.L., Moyer, E.J., Srivier, R.L.: Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression. *J. Clim.* **31**(20), 8573–8588 (2018)
- Huang, H., Chen, Z.: Bayesian composite quantile regression. *J. Stat. Comput. Simul.* **85**(18), 3744–3754 (2015)
- Jiang, R., Zhou, Z.G., Qian, W.M., Chen, Y.: Two step composite quantile regression for single-index models. *Comput. Stat. Data Anal.* **64**, 180–191 (2013)
- Jiang, X., Jiang, J., Song, X.: Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica*, p. 1479–1506 (2012)
- Kai, B., Li, R., Zou, H.: Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *J. R. Stat. Soc. Ser. B Stat Methodol.* **72**(1), 49–69 (2010)
- Kindo, B.P.: Bayesian Ensemble of Regression Trees for Multinomial Probit and Quantile Regression. Phd thesis, University of South Carolina (2016)
- Kindo, B.P., Wang, H., Hanson, T., Peña, E.A.: Bayesian quantile additive regression trees (2016). <https://arxiv.org/abs/1607.02676>, [arXiv:1607.02676](https://arxiv.org/abs/1607.02676) [stat.ML]
- Koenker, R.: *Quantile regression*, vol. 38. Cambridge University Press, Cambridge (2005)
- Koenker, R., Bassett Jr, G.: Regression quantiles. *Econometrica: journal of the Econometric Society*, p. 33–50 (1978)
- Koenker, R., Hallock, K.F.: Quantile regression. *J. Econ. Perspect.* **15**(4), 143–156 (2001)
- Koenker, R., Ng, P., Portnoy, S.: Quantile smoothing splines. *Biometrika* **81**(4), 673–680 (1994)
- Kozumi, H., Kobayashi, G.: Gibbs sampling methods for Bayesian quantile regression. *J. Stat. Comput. Simul.* **81**(11), 1565–1578 (2011)
- Lee, D., Neocleous, T.: Bayesian quantile regression for count data with application to environmental epidemiology. *J. R. Stat. Soc. Ser. C: Appl. Stat.* **59**(5), 905–920 (2010)
- Lee, J., Hwang, B.S.: Ordered probit Bayesian additive regression trees for ordinal data. *Stat.* **13**(1), e643 (2024)
- Li, Q., Xi, R., Lin, N.: Bayesian Regularized Quantile Regression. *Bayesian Anal.* **5**(3), 533–556 (2010)
- Linero, A.R.: Bayesian regression trees for high-dimensional prediction and variable selection. *J. Am. Stat. Assoc.* **113**(522), 626–636 (2018)
- Linero, A.R., Yang, Y.: Bayesian regression tree ensembles that adapt to smoothness and sparsity. *J. R. Stat. Soc. Ser. B Stat Methodol.* **80**(5), 1087–1110 (2018)
- Marrocu, E., Paci, R., Zara, A.: Micro-economic determinants of tourist expenditure: a quantile regression approach. *Tour. Manage.* **50**, 13–30 (2015)
- Powell, D.: Quantile regression with nonadditive fixed effects. *Empirical Economics* **63**(5), 2675–2691 (2022)
- Qian, W., Yang, Y., Zou, H.: HDtweedie: The Lasso for Tweedie's Compound Poisson Model Using an IRLS-BMD Algorithm (2022), <https://CRAN.R-project.org/package=HDtweedie>, r package version 1.2
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2024), <https://www.R-project.org/>
- Reich, B.J., Fuentes, M., Dunson, D.B.: Bayesian spatial quantile regression. *J. Am. Stat. Assoc.* **106**(493), 6–20 (2011)
- Sparapani, R.A., Logan, B.R., McCulloch, R.E., Laud, P.W.: Nonparametric survival analysis using Bayesian additive regression trees (BART). *Stat. Med.* **35**(16), 2741–2753 (2016)
- Sriram, K., Ramamoorthi, R., Ghosh, P.: Posterior Consistency of Bayesian Quantile Regression Based on the Misspecified Asymmetric Laplace Density. *Bayesian Anal.* **8**(2), 479–504 (2013)
- Tan, Y.V., Roy, J.: Bayesian additive regression trees and the General BART model. *Stat. Med.* **38**(25), 5048–5069 (2019)
- Taylor, J.W.: A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *J. Forecast.* **19**(4), 299–311 (2000)
- Thompson, P., Cai, Y., Moyeed, R., Reeve, D., Stander, J.: Bayesian nonparametric quantile regression using splines. *Comput. Stat. Data Anal.* **54**(4), 1138–1150 (2010)
- Um, S., Linero, A.R., Sinha, D., Bandyopadhyay, D.: Bayesian additive regression trees for multivariate skewed responses. *Stat. Med.* **42**(3), 246–263 (2023)
- Venables, W.N., Ripley, B.D.: *Modern Applied Statistics with S*. Fourth ed. New York: Springer; (2002). <https://www.stats.ox.ac.uk/pub/MASS4/>, ISBN 0-387-95457-0

- Wei, Y., Kehm, R.D., Goldberg, M., Terry, M.B.: Applications for quantile regression in epidemiology. *Curr. Epidemiol. Rep.* **6**, 191–199 (2019)
- Wei, Y., Pere, A., Koenker, R., He, X.: Quantile regression methods for reference growth charts. *Stat. Med.* **25**(8), 1369–1382 (2006)
- Xu, Q., Deng, K., Jiang, C., Sun, F., Huang, X.: Composite quantile regression neural network with applications. *Expert Syst. Appl.* **76**, 129–139 (2017)
- Xu, S.G., Reich, B.J.: Bayesian nonparametric quantile process regression and estimation of marginal quantile effects. *Biometrics* **79**(1), 151–164 (2023)
- Xu, Y., Hogan, J., Daniels, M., Kantor, R., I., Mwangi, A.: Augmentation Samplers for Multinomial Probit Bayesian Additive Regression Trees. *Journal of Computational and Graphical Statistics*, **34**(2), 498–508 (2024)
- Yang, Y., Wang, H.J., He, X.: Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *Int. Stat. Rev.* **84**(3), 327–344 (2016)
- Yu, K., Jones, M.: Local linear quantile regression. *J. Am. Stat. Assoc.* **93**(441), 228–237 (1998)
- Yu, K., Moyeed, R.A.: Bayesian quantile regression. *Stat. Probab. Lett.* **54**(4), 437–447 (2001)
- Yuan, X., Xiang, X., Zhang, X.: Bayesian composite quantile regression for the single-index model. *PLoS ONE* **18**(5), e0285277 (2023)
- Zhang, T., Geng, G., Liu, Y., Chang, H.H.: Application of Bayesian additive regression trees for estimating daily concentrations of PM_{2.5} components. *Atmosphere* **11**(11), 1233 (2020)
- Zhao, K., Lian, H.: A note on the efficiency of composite quantile regression. *J. Stat. Comput. Simul.* **86**(7), 1334–1341 (2016)
- Zhao, Z., Xiao, Z.: Efficient regressions via optimally combining quantile information. *Economet. Theor.* **30**(6), 1272–1314 (2014)
- Zheng, Q., Peng, L., He, X.: Globally adaptive quantile regression with ultra-high dimensional data. *Ann. Stat.* **43**(5), 2225 (2015)
- Zou, H., Yuan, M.: Composite quantile regression and the oracle model selection theory. *Ann. Stat.* **36**(3), 1108–1126 (2008)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.