Bayesian Relative Composite Quantile Regression Approach of Ordinal Latent Regression Model with $L_{1/2}$ Regularization

Tian Yu-Zhu[1]*, Wu Chun-Ho[2], Tai Ling-Nan[3], Mian Zhi-Bao[4], Tian Mao-Zai[5]

[1]School of Mathematics and Statistics, Northwest Normal University, Lanzhou, China

[2]School of Decision Science, The Hang Seng University of Hong Kong, Hong Kong

[3]School of Economics and Management, The Open University of China, Beijing, China

[4]Department of Computer Science, University of Hull, United Kingdom

[5]School of Statistics, Renmin University of China, Beijing, China

March 22, 2024

Abstract

Ordinal data frequently occur in various fields such as knowledge level assessment, credit rating, clinical disease diagnosis, and psychological evaluation. The classic models including cumulative logistic regression or probit regression are often used to model such ordinal data. But these modeling approaches conditionally depict the mean characteristic of response variable on a cluster of predictive variables, which often result in non-robust estimation results. As a considerable alternative, composite quantile regression (CQR) approach is usually employed to gain more robust and relatively efficient results. In this paper, we propose a Bayesian CQR modeling approach for ordinal latent regression model. In order to overcome the recognizability problem of the considered model and obtain more robust estimation results, we advocate to using the Bayesian relative CQR approach to estimate regression parameters. Additionally, in regression modeling, it is a highly desirable task to obtain a parsimonious model that retains only important covariates. We incorporate the Bayesian $L_{1/2}$ penalty into the ordinal latent CQR regression model to simultaneously conduct parameter estimation and variable selection. Finally, the proposed Bayesian relative CQR approach is illustrated by Monte Carlo simulations and a real data application. Simulation results and real data example show that the suggested Bayesian relative CQR approach has good performance for the ordinal regression models.

Keywords: Latent regression model; Ordinal response; Monte Carlo; CQR modeling; $L_{1/2}$ penalty

1 Introduction

In regression modeling, we often encounter response variables whose data type is ordinal. For example, in the analysis of factors affecting credit rating, the responded credit rating

*The corresponding author is Tian Yuzhu, pole1999@163.com.

is typically ordinal; in the analysis of risk factors affecting disease development, the degree of disease can also be considered as ordinal; in the analysis of students' knowledge level, the knowledge ability is classified into several ordinal levels. Ordinal regression models including logistic regression and probit regression are generally employed to fit this type of data. Another usual alternative is ordinal latent regression which is also used to capture the mean characteristic through the continuous latent responses. One can refer to Cliff (1996), Zhang *et al.*(2003), Agresti (2010), Manuguerra & Heller (2010), Montesinos-lopez *et al.* (2015), Sha & Dechi (2019) and Tutz (2022) for a more detailed discussion. However, the above modeling approaches mainly capture the mean characteristic of response variable conditionally on covariates, which may result in non-robust estimates in the presence of outliers. Unlike only modeling the empirical means, the QR approach studies the full conditional distributions of response variable. One can refer to Davino *et al.*(2014) for a comprehensive summary of QR modeling. However, constructing QR estimates for ordinal variables becomes challenging because quantiles of ordinal categorical data cannot be obtained through a simple inverse operation of the cumulative distribution function (CDF). Meanwhile, the standard QR depicts a conditional distribution of the dependent variable for a single quantile, making it challenging to select the most informative quantile to gain efficient estimators.

As an alternative, the CQR approach combines the information of multiple quantile levels which produce more robust and efficient estimation results than the mean regression and single QR modeling. There have been many existing literatures on CQR modeling in the last fifteen years. One can refer to Zou & Yuan (2008), Kai *et al.* (2010), Tang *et al.*(2012), Jiang *et al.* (2014), Wang *et al.* (2018), Huang & Zhan (2022), etc. Additionally, from the Bayesian framework, Huang & Chen (2015) and Alhamzawi (2016) discussed the Bayesian CQR of linear models, Tian *et al.* (2017) and Tian *et al.* (2021) studied Bayesian CQR for linear mixed models and weighted CQR of longitudinal data using MCEM algorithm, respectively. In this paper, we will focus on the Bayesian CQR modeling of the ordinal latent regressions.

In high-dimensional regression modeling, many covariates are usually included in the model, but only a small part are statistically significant. Many regularization methods

including LASSO (Tibshirani (1996): Least absolute shrinkage and select operator), adaptive LASSO (Zou (2006)), SCAD (Fan & Li (2001): Smoothly clipped absolute deviation), Enet (Hui & Hastie (2005): the Elastic-net), bridge penalized regression (Fu (1998)), and $L_{1/2}$-norm penalization (Xu (2010)) can be frequently used to conduct variable selection and model estimation simultaneously. In the Bayesian framework, Park & Casella (2008) proposed Bayesian LASSO regression, Alhamzawi *et al.* (2012) considered Bayesian adaptive LASSO QR, Polson *et al.* (2014) studied Bayesian bridge regression, Betancourt *et al.* (2017) studied Bayesian fused Lasso regression for dynamic binary networks, Alhamzawi & Ali (2018) discussed Bayesian $L_{1/2}$ Tobit QR, Mallick & Yi (2018) considered Bayesian $L_{1/2}$ regularization. Based on the above literature review, $L_{1/2}$ penalty will be incorporated into the Bayesian CQR modeling for the given ordinal model.

The remainder of this paper is organized as follows. Section 2 introduces the latent ordinal regression model and the working likelihood. Section 3 presents the Bayesian algorithm of the considered method. The selection of CQR level K and the relative CQR estimation approach are also highlighted in the Subsection 3.4. Section 4 provides some Monte Carlo simulations to illustrate the proposed modeling approach. Section 5 presents two real-world applications to illustrate the proposed estimation procedure. The last section draws some conclusions.

2 The ordinal latent CQR model and the working likelihood

2.1 The latent CQR model

The ordinal responses $y_i, i = 1, \cdots, N$ are linked on the latent unobservable responses y_i^* as follows

$$y_i = \begin{cases} 1, & \delta_0 < y_i^* \leq \delta_1; \\ r, & \delta_{r-1} < y_i^* \leq \delta_r; \quad r = 2, \cdots, R-1, \\ R, & \delta_{R-1} < y_i^* \leq \delta_R; \end{cases} \tag{2.1}$$

where $\delta_0, \cdots, \delta_R$ are cut-points whose coordinates satisfy $-\infty = \delta_0 < \delta_1 < \cdots < \delta_{R-1} < \delta_R = +\infty$, δ_{r-1} and δ_r define the lower and upper thresholds of the interval corresponding to observed outcome r. The latent responses y_i^* are assumed to be generated from the

following linear model

$$y_i^* = x_i^T \beta + \varepsilon_i, \tag{2.2}$$

where $\beta = (\beta_1, \cdots, \beta_p)^T$ is the vector of regression coefficients, $x_i = (x_{i1}, \cdots, x_{ip})^T$ is the explaining variable, ε_i is the random error term.

For latent regression model (2.2), CQR estimator $\hat{\beta}^{CQR}$ can be derived by minimizing the objective loss function

$$(\hat{\alpha}_1, \cdots, \hat{\alpha}_K, \hat{\beta}^{CQR}) = \arg \min_{\alpha_1, \cdots, \alpha_K, \beta} \sum_{i=1}^{N} \sum_{k=1}^{K} \rho_{\tau_k}(y_i^* - x_i^T \beta - \alpha_k), \tag{2.3}$$

where α_k is the τ_k-th quantile of the error term ε_i and satisfy monotonicity: $\alpha_1 < \cdots < \alpha_K$. The composite quantile levels can be simply set to $\tau_k = \frac{k}{K+1}, k = 1, \cdots, K$. Evidently, the CQR for the case of $K = 1$ will degenerate into median regression. Although the CQR estimator can result in higher estimation efficiency, the computation is a challenging work due to the complexity of the objective function (2.3). Tian *et al.* (2016) studied a likelihood-based CQR approach by using the CALD (Composite asymmetric Laplace distribution) and derived an IWLSE (Iterative weighted least square estimation) solution.

Borrowing the CALD in Tian *et al.* (2016), the CQR working likelihood of y_i^* in model (2.2) can be represented by

$$\prod_{i=1}^{N} f(y_i^* | \mu_i, \sigma) \propto \prod_{i=1}^{N} \prod_{k=1}^{K} \frac{1}{\sigma} \exp\Big\{ - \rho_{\tau_k}\Big(\frac{y_i^* - \mu_{ik}}{\sigma}\Big)\Big\}, \tag{2.4}$$

where $\mu_{ik} = x_i^T \beta + \alpha_k$ is the τ_k conditional quantile of the response y_i^*. In order to carry out the fully Bayesian inference, the CQR working likelihood (2.4) can be represented as the following hierarchical likelihood

$$\begin{cases} \prod_{i=1}^{N} f(y_i^* | \upsilon_i, \mu_i, \sigma) \propto \prod_{i=1}^{N} \prod_{k=1}^{K} \frac{1}{\sqrt{\theta_{2,k} \sigma \upsilon_{ik}}} \exp\Big\{ - \frac{(y_i^* - \mu_{ik} - \theta_{1,k} \upsilon_{ik})^2}{2\sigma\theta_{2,k}\upsilon_{ik}}\Big\}, \\ \upsilon_{ik} \sim \text{Exp}(\frac{1}{\sigma}), \end{cases} \tag{2.5}$$

where $\theta_{1,k} = \frac{1-2\tau_k}{\tau_k(1-\tau_k)}$, $\theta_{2,k} = \frac{2}{\tau_k(1-\tau_k)}$, $\mu_i = (\mu_{i1}, \cdots, \mu_{iK})$, and $\upsilon_i = (\upsilon_{i1}, \cdots, \upsilon_{iK})$ is the latent variable, $\text{Exp}(\frac{1}{\sigma})$ denotes the exponential distribution with parameter $\frac{1}{\sigma}$.

Based on the hierarchical likelihood (2.5), we derive the conditional posterior distri-

bution of latent variable y_i^* as follows

$$\begin{aligned} f(y_i^*|\upsilon_i,\mu_i,\sigma) &\propto \prod_{k=1}^{K}\frac{1}{\sqrt{\sigma\theta_{2,k}\upsilon_{ik}}}\exp\Big\{-\frac{(y_i^*-\mu_{ik}-\theta_{1,k}\upsilon_{ik})^2}{2\theta_{2,k}\sigma\upsilon_{ik}}\Big\} \\ &\propto \exp\Big\{-\frac{1}{2}\sum_{k=1}^{K}\frac{(y_i^*-\mu_{ik}-\theta_{1,k}\upsilon_{ik})^2}{\sigma\theta_{2,k}\upsilon_{ik}}\Big\} \\ &\propto \exp\Big\{-\frac{1}{2}\Big[y_i^{*2}\cdot\sum_{k=1}^{K}\frac{1}{\sigma\theta_{2,k}\upsilon_{ik}}-2y_i^*\cdot\sum_{k=1}^{K}\frac{\mu_{ik}+\theta_{1,k}\upsilon_{ik}}{\sigma\theta_{2,k}\upsilon_{ik}}\Big]\Big\} \\ &\propto \exp\Big\{-\frac{1}{2}\sum_{k=1}^{K}\frac{1}{\sigma\theta_{2,k}\upsilon_{ik}}\cdot\Big(y_i^*-\frac{\sum_{k=1}^{K}\frac{\mu_{ik}+\theta_{1,k}\upsilon_{ik}}{\theta_{2,k}\upsilon_{ik}}}{\sum_{k=1}^{K}\frac{1}{\theta_{2,k}\upsilon_{ik}}}\Big)^2\Big\} \\ &\sim N\Big(\frac{\sum_{k=1}^{K}\frac{\mu_{ik}+\theta_{1,k}\upsilon_{ik}}{\theta_{2,k}\upsilon_{ik}}}{\sum_{k=1}^{K}\frac{1}{\theta_{2,k}\upsilon_{ik}}},\frac{\sigma}{\sum_{k=1}^{K}\frac{1}{\theta_{2,k}\upsilon_{ik}}}\Big) \\ &\triangleq N(u_i,\varphi_i^2). \end{aligned} \tag{2.6}$$

The conditional distribution $f(y_i^*|\upsilon_i,\mu_i,\sigma)$ will be used to conduct Bayesian posterior inference of latent variable y_i^* in Subsection 3.2.

2.2 The joint likelihood

Denote $y=(y_1,\cdots,y_N)$, $y^*=(y_1^*,\cdots,y_N^*)$,$\upsilon=\{\upsilon_1,\cdots,\upsilon_N\}$, $\alpha=(\alpha_1,\cdots,\alpha_K)$ and $x=\{x_1,\cdots,x_N\}$. Based on (2.6) and model (2.1), the conditional cumulative distribution function (CDF) of observed response y_i is

$$\begin{aligned} P(y_i\le r|y_i^*,x_i,\beta,\sigma,\upsilon,\delta_r,\alpha) &= P(y_i^*\le\delta_r|x_i,\upsilon_i,\beta,\sigma,\alpha) \\ &= F_{y_i^*}(\delta_r)=\Phi(\frac{\delta_r-u_i}{\varphi_i}), \end{aligned} \tag{2.7}$$

where Φ is the CDF of standard normal distribution. Then, the probability of y_i being at r-th category is

$$\begin{aligned} \pi_{ir} &= P(y_i=r|y_i^*,x_i,\beta,\sigma,\upsilon_i,\delta_{r-1},\delta_r,\alpha) \\ &= P(\delta_{r-1}<y_i^*\le\delta_r|x_i,\upsilon_i,\beta,\sigma,\alpha) \\ &= \Phi(\frac{\delta_r-u_i}{\varphi_i})-\Phi(\frac{\delta_{r-1}-u_i}{\varphi_i}). \end{aligned} \tag{2.8}$$

Hence, the conditional likelihood of observation data y can be expressed as

$$P(y|y^*,x,\upsilon,\beta,\sigma,\delta,\alpha)=\prod_{i=1}^{N}\prod_{r=1}^{R}\pi_{ir}^{I_{ir}}, \tag{2.9}$$

where δ=$\{\delta_1, \cdots, \delta_{R-1}\}$, I_{ir} is the indictor function of y_i, valued as 1 when $y_i = r$, or 0 otherwise.

The marginal likelihood of the CQR for ordinal response y is

$$\begin{aligned} L_O(\Theta|y,x) &= \int \prod_{i=1}^{N} \Big[P(y_i|y_i^*, x_i, \upsilon_i, \beta, \sigma, \delta, \alpha) \cdot f(y_i^*|x_i, \upsilon_i, \beta, \sigma, \alpha) \\ &\quad \times \prod_{k=1}^{K} f(\upsilon_{ik}|\sigma_k) \Big] d\upsilon dy^*, \end{aligned} \tag{2.10}$$

where Θ is a set of unknown parameters, $f(y_i^*|x_i, \upsilon_i, \beta, \sigma, \alpha)$ is the conditional normal pdf (2.6), $f(\upsilon_{ik}|\sigma)$ is the pdf of exponential distribution in (2.5). The marginal likelihood (2.10) encounters tedious high-dimensional integral which is difficult to maximize. MCMC (Markov Chain Monte Carlo) algorithm can be naturally used to address the computational problem. To conduct the fully Bayesian inference in Section 3, the joint hierarchical likelihood of the complete data $\{y, y^*, v\}$ can be derived as follows

$$\begin{aligned} L_C(\Theta|y,y^*,v) &= \prod_{i=1}^{N} \left[P(y_i|y_i^*, x_i, \upsilon_i, \beta, \sigma, \delta, \alpha) \cdot f(y_i^*|x_i, \upsilon_i, \beta, \sigma, \alpha) \times \prod_{k=1}^{K} f(\upsilon_{ik}|\sigma) \right] \\ &= P(y|y^*, x, \upsilon, \beta, \sigma, \delta, \alpha) \cdot f(y^*|x, \upsilon, \beta, \sigma, \alpha) \cdot f(\upsilon|\sigma). \end{aligned} \tag{2.11}$$

3 Bayesian procedures

3.1 Speciation of priors

Firstly, for regression coefficient β, we impose penalization priors to conduct variable selection. Commonly used penalty functions mainly include L_1-norm penalty, L_2-norm penalty, SCAD penalty, L_ξ $(0 < \xi < 1)$ penalty. Xu (2010) revealed that $L_{1/2}$ penalty is the most sparse and robust among the L_ξ. For achieving $L_{1/2}$ penalized CQR for the considered model, we exert a generalized Gaussian distribution (GGD) prior on regression coefficient β as follows

$$\pi(\beta|\sigma,\lambda) = \prod_{j=1}^{p} \pi(\beta_j|\sigma,\lambda), \ \pi(\beta_j|\sigma,\lambda) = \frac{(\frac{\lambda}{\sigma})^2}{2 \cdot \Gamma(3)} \exp\Big\{ -\frac{\lambda}{\sigma}|\beta_j|^{1/2} \Big\}, \tag{3.1}$$

where $\lambda > 0$ is penalty parameter.

Mallick & Yi (2018) proposed a hierarchical representation of the GGD, which is used to conduct more efficient Bayesian inference. Based on such a hierarchical form, the

prior of β_j can be represented by

$$\begin{cases} \pi(\beta_j|\sigma,\lambda) = \int_0^\infty \pi(\beta_j|s_j)\cdot\pi(s_j|\lambda,\sigma)ds_j, \\ \pi(\beta_j|s_j) = \text{Uniform}(-s_j^2, s_j^2), \pi(s_j|\lambda,\sigma) = \text{Gamma}\Big(3, \frac{\lambda}{\sigma}\Big). \end{cases}$$

Thereupon, the prior of β is hierarchically expressed as

$$\pi(\beta|\sigma,\lambda) \propto \prod_{j=1}^{p} \big[\pi(\beta_j|\lambda,\sigma,s_j)\cdot\pi(s_j|\lambda,\sigma)\big]. \tag{3.2}$$

For threshold parameter $\delta = (\delta_1, \cdots, \delta_{R-1})$, the prior is selected according to the recommendation of Sha & Dechi (2019). This method has good identifiability for the monotonicity of threshold components. Suppose $G(\cdot)$ is the CDF of one continuous variable whose domain lies in $(-\infty, \infty)$, such that the probability of falling on each interval are $p_r = P(\delta_{r-1} < \Delta < \delta_r) = G(\delta_r) - G(\delta_{r-1})$, $j = 1, \cdots, R$ and satisfy $\sum_{r=1}^{R} p_r = 1$. By mathematical transformation, we derive

$$\begin{cases} \delta_1 = G^{-1}(p_1), \\ \delta_2 = G^{-1}(p_1 + p_2), \\ \vdots \\ \delta_{R-1} = G^{-1}(p_1 + \cdots + p_{_{R-1}}). \end{cases}$$

For the probability parameters $p_1, \cdots, p_{_R}$, the priors are commonly set as the Dirichlet distribution $\pi(p_1, \cdots, p_{_R}|\gamma) = \frac{1}{B(\gamma)}\prod_{r=1}^{R} p_r^{\gamma_r - 1}$, where $\gamma = (\gamma_1, \cdots, \gamma_R)$ is positive hyperparameter vector, $B(\gamma) = \prod_{r=1}^{R}\Gamma(\gamma_r)/\Gamma(\sum_{r=1}^{R}\gamma_r)$, $\Gamma(\cdot)$ denotes the gamma function. By using the above transformation, the prior of δ becomes

$$\pi(\delta) = \frac{1}{B(\gamma)}\prod_{r=1}^{R}[G(\delta_r) - G(\delta_{r-1})]^{\gamma_r - 1}\cdot\prod_{r=1}^{R-1} g(\delta_r).$$

where $g(\cdot)$ is the density function of $G(\cdot)$.

The parameter λ is selected as the conjugate Gamma prior: $\pi(\lambda) = \text{Gamma}(a, b)$.

The scale σ is set as the conjugate inverse Gamma prior: $\pi(\sigma) = \text{IGamma}(c, d)$.

For given K, the prior of α is simply taken as $\pi(\alpha) = \prod_{k=1}^{K}\pi(\alpha_k)$, where $\pi(\alpha_k) \sim N(\alpha_{k,0}, \varsigma_{k,0}^2)$, $\alpha_{k,0}$ and $\varsigma_{k,0}^2$ are mean and variance hyperparameters. Quantile parameters $\alpha_1, \cdots, \alpha_K$ need to satisfy the constraint: $\alpha_1 < \cdots < \alpha_K$. For retaining the monotonicity of their posterior estimates, the hyperparameters $\alpha_{k,0}, k = 1, \cdots, K$ can be set as $\alpha_{1,0} <$

$\alpha_{2,0} < \cdots < \alpha_{K,0}$. The scale hyperparameters $\varsigma_{k,0}^2$ can be taken as the same value for convenience. A rational remark concerning the hyperparameter setting of prior $\pi(\alpha)$ can be found in Tian *et al.* (2019).

The joint prior of all unknown parameters can be expressed as follows,

$$\pi(\beta|\sigma,\lambda)\pi(\lambda)\pi(\sigma)\pi(\delta)\pi(\alpha). \tag{3.3}$$

3.2 The joint posterior

Incorporating the joint prior (3.3) into the joint likelihood (2.11) results in the joint posterior density of parameters and latent variables as follows

$$P(\beta,\sigma,\lambda,\delta,\alpha,\upsilon,y^*|y,x) \propto L_C(\Theta|y,y^*,x,\upsilon)\cdot\pi(\beta|\sigma,\lambda)\pi(\lambda)\pi(\sigma)\pi(\delta)\pi(\alpha). \tag{3.4}$$

The hierarchical representation of joint posterior density (3.4) is

$$\begin{cases} y|y^*,x,\beta,\sigma,\delta,\upsilon \sim P(y|y^*,x,\beta,\sigma,\delta,\upsilon,\alpha), \\ y^*|x,\beta,\sigma,\upsilon \sim f(y^*|x,\beta,\sigma,\upsilon,\alpha), \\ \upsilon_{ik}|\sigma \sim Exp\Big(\frac{1}{\sigma}\Big), i=1,\cdots,N, k=1,\cdots,K; \\ \beta|S \sim \prod_{j=1}^{p} \text{Uniform}(-s_j^2, s_j^2), \\ S|\lambda,\sigma \sim \prod_{j=1}^{p} \text{Gamma}\Big(3,\frac{\lambda}{\sigma}\Big), \\ \lambda \sim \text{Gamma}(a,b), \\ \sigma \sim \text{IGamma}(c,d), \\ \alpha \sim \prod_{k=1}^{K}\pi(\alpha_k) \\ \delta \sim \pi(\delta). \end{cases} \tag{3.5}$$

The above joint posterior (3.4) is quite high-dimensional and complex to calculate its posterior quantities. MCMC algorithm is employed to derive the posterior samples for conducting Bayesian inference.

3.3 Gibbs sampler algorithm

by using the Gibbs sampler procedure of MCMC algorithm, we derive the fully conditional posterior distributions of parameters and latent variables as follows.

$\circ$ Sample δ from its fully conditional posterior distribution $\pi(\delta)\cdot\prod_{r=1}^{R}\prod_{i=1}^{N} I(\delta_{r-1} < y_i^* \le \delta_r, y_i = r)$. Specially, the coordinate δ_r can be sampled conditionally on $\delta_{(-r)} =$

$(\delta_1, \cdots, \delta_{r-1}, \delta_{r+1}, \cdots, \delta_{R-1})$ for $r = 1, \cdots, R-1$. It can be noticed that the fully conditional posterior distribution of δ_r is

$$\pi(\delta_r | y, \delta_{(-r)}) \propto [G(\delta_{r+1}) - G(\delta_r)]^{\gamma_{r+1}-1} \cdot [G(\delta_r) - G(\delta_{r-1})]^{\gamma_r - 1} \cdot g(\delta_r) \cdot I(\omega_{r1} < \delta_r < \omega_{r2}),$$

where $\omega_{r1} = max(y_i^* | y_i = r), \omega_{r2} = min(y_i^* | y_i = r+1)$ and $G(\delta_0) = 0, G(\delta_R) = 1$. In applications, the distribution function $G(\cdot)$ can be simply set as a normal distribution with mean μ_0 and large scale σ_0 for covering a wide range in $(-\infty, \infty)$, where μ_0 and σ_0 are hyperparameters. Conditionally, δ_r is a random variable whose transformation $G(\delta_r)$ is distributed as the shifted $G(\delta_{r-1})$ and scaled $G(\delta_{r+1}) - G(\delta_{r-1})$ Beta distribution truncated at the interval $[G(\omega_{r1}), G(\omega_{r2})]$, equivalently,

$$\frac{G(\delta_r) - G(\delta_{r-1})}{G(\delta_{r+1}) - G(\delta_{r-1})} \cdot I(\delta_r \in \delta_{r-1}, \delta_{r+1}) \sim Beta(\gamma_r, \gamma_{r+1})$$

truncated at $\Big[\frac{G(\omega_{r1}) - G(\delta_{r-1})}{G(\delta_{r+1}) - G(\delta_{r-1})}, \frac{G(\omega_{r2}) - G(\delta_{r-1})}{G(\delta_{r+1}) - G(\delta_{r-1})}\Big]$. Specially, first draw a η_r from the above truncated Beta distribution and then get $\delta_r = G^{-1}(\xi_r)$, where $\xi_r = G(\delta_{r-1}) + \eta_r \cdot (G(\delta_{r+1}) - G(\delta_{r-1}))$, $r = 1, \cdots, R-1$.

- Sample υ_{ik} from the generalized inverse Gaussian distribution

$$\text{GIG}\Big(\frac{1}{2},\ \frac{\iota_{ik}^2}{\theta_{2,k}\sigma},\ \frac{\theta_{1,k}^2 + 2\theta_{2,k}}{\theta_{2,k}\sigma}\Big),$$

where $\iota_{ik} = y_i^* - \alpha_k - x_i^T \beta$.

- Sample σ from the inverse Gamma distribution

$$\text{IG}\Big(\frac{3NK}{2} + 3p + c,\ \sum_{k=1}^{K}\sum_{i=1}^{N}\Big(\frac{e_{ik}^2}{2\theta_{2,k}\upsilon_{ik}} + \upsilon_{ik}\Big) + \lambda\sum_{j=1}^{p} s_j + d\Big),$$

where $e_{ik} = y_i^* - \alpha_k - x_i^T\beta - \theta_{1,k}\upsilon_{ik}$.

- Sample β from the multivariate truncated normal distribution

$$N(\beta^*,\ B^*) \cdot \prod_{j=1}^{p} \text{I}(|\beta_j| < s_j^2),$$

where$\beta^* = B^* \cdot \Big(\frac{1}{\sigma}\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{x_i \cdot (y_i^* - \alpha_k - \theta_{1,k}\upsilon_{ik})}{\theta_{2,k}\upsilon_{ik}}\Big)$, $B^* = \Big(\frac{1}{\sigma}\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{x_i x_i^T}{\theta_{2,k}\upsilon_{ik}}\Big)^{-1}$.

Specially, the coordinates $\beta_j, j = 1, \cdots, p$ of β can be conditionally sampled from the following truncated normal distribution

$$TN_{(-s_j^2, s_j^2)}(\mu_{\beta_j}, \sigma_{\beta_j}^2),$$

where $\sigma^2_{\beta_j} = \left(\frac{1}{\sigma}\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{x_{ij}^2}{\theta_{2,k}\upsilon_{ik}}\right)^{-1}, \mu_{\beta_j} = \sigma^2_{\beta_j}\cdot\left(\frac{1}{\sigma}\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{x_{ij}\eta_{ik}}{\theta_{2,k}\upsilon_{ik}}\right)$, $\eta_{ik} = y_i^* - c_k - \sum_{u\neq j} x_{iu}\beta_u - \theta_{1,k}\upsilon_{ik}$.

∘ Sample $s_j, j = 1, \cdots, p$ from the left-truncated exponential distribution $\text{Exp}(\lambda/\sigma)I\{s_j > |\beta_j|^{1/2}\}$, using the inversion method, which can be enforced by two steps: (I) Sample $s_j^* \sim \text{Exp}(\lambda/\sigma)$; (II) Compute $s_j = s_j^* + |\beta_j|^{1/2}$.

∘ Sample λ from the Gamma distribution $\text{Gamma}(3p + a,\ b + \sum_{j=1}^{p} s_j/\sigma)$.

∘ Sample $\alpha_k, k = 1, \cdots, K$ from the normal distribution $N\Big(\alpha^*_{k,0},\ (\varsigma^2_{k,0})^*\Big)$, where $\alpha^*_{k,0} = (\varsigma^2_{k,0})^*\left(\frac{1}{\sigma}\sum_{i=1}^{N}\frac{\epsilon_{ik}}{\theta_{2,k}\upsilon_{ik}} + \frac{\alpha_{k,0}}{\varsigma^2_{k,0}}\right)$, $(\varsigma^2_{k,0})^* = \left(\frac{1}{\sigma}\sum_{i=1}^{N}\frac{1}{\theta_{2,k}\upsilon_{ik}} + \frac{1}{\varsigma^2_{k,0}}\right)^{-1}, \epsilon_{ik} = y_i^* - x_i^T\beta - \theta_{1,k}\upsilon_{ik}$.

∘ Sample $y_i^*, i = 1, \cdots, N$ from the truncated normal distribution

$$TN_{(\delta_{r-1},\delta_r)}(u_i, \varphi_i^2), y_i = r, r = 1, \cdots, R.$$

3.4 Selecting K and the relative CQR estimation

The selection of K is very critical in CQR applications. Although employing bigger K in CQR modeling can produce higher estimation efficiency, more computational burden encounters. A optimal K is an equilibrium between estimation efficiency and model complexity. Many authors have found that the efficiency gain is relatively insignificant as K increases. Jiang *et al.* (2014) studied CQR with $K = 7$ for the DTARCH model, and they showed that increasing K does not change the results significantly. Huang & Chen (2015) studied Bayesian CQR with $K = 9$ and declared that they tried several other values of K from 5 to 20 and found the numerical results are not sensitive to this choice. In addition, Tian *et al.* (2016, 2017) discussed CQR with the value of K from 3 to 9 and found that increasing the bigger value for K does not results in higher efficiency gain significantly. Hence, the value of K in CQR analysis is reasonably recommended to take from 3 to 9.

Another important issue in this paper is identifiability of ordinal latent regression model. Recently, Grabski et al. (2019) proposed a relative estimation approach for the ordinal QR model to allow adaptive cutpoints for yielding identifiable results. The inference addressed the ratios of the coefficients to the cutpoint vector, which is identifiable, rather

than on the magnitudes of the original coefficients. We employ such a proposal to define the Bayesian relative CQR estimators for ordinal latent regression coefficients. In simulations and real data analysis, Bayesian relatively CQR estimates and standard deviations (std) of ratio parameters $\frac{\beta_j}{\delta_{R-1}}, j = 1, \cdots, p$ are reported by using the corresponding posterior samples, instead of original parameters β_j. Then, to specify the statistical significance of the coefficients based on the CQR approach, we use $\frac{\hat{\beta}_j}{\hat{\delta}_{R-1}}$ to affirm whether each β_j is significantly different from 0 or not. Meanwhile, in Section 5, for the aim of comparison, we derive the restored CQR estimates $\hat{\beta}_j$ of original coefficients by multiplying the relative CQR estimates $\frac{\hat{\beta}_j}{\hat{\delta}_{R-1}}$ by the estimated $\hat{\delta}_{R-1}$.

4 Simulation studies

4.1 Model parameters and data generation

In this section, simulation studies are presented to illustrate the sample performance of the proposed Bayesian relative CQR approach. We generate 100 datasets from the latent regression model (2.2) with $N = 200$, where covariates x_i are generated from the multivariate standard normal distribution and error ε_i are generated from two cases: (1) standard normal distribution ($N(0, 1)$); (2) student t distribution with three degree (t_3). Consider the following two cases for regression coefficient vector.

Model 1: Sparse case with $\beta = (1.2, 1.2, 0, 0, 0, 0)^T$,

Model 2: High-dimensional sparse case with $\beta = (\beta_{1:2}^T, \beta_{3:20}^T)^T$, where $\beta_{1:2} = (1.2, 1.2)^T, \beta_{3:20} = (0, \cdots, 0)^T$.

The latent responses y_i^* are divided into four ordinal categories based on the given thresholds which result in the observed responses y_i as follows

$$y_i = \begin{cases} 1, & -\infty < y_i^* \leq \delta_1, \\ 2, & 0 < y_i^* \leq \delta_2, \\ 3, & 1.6 < y_i^* \leq \delta_3, \\ 4, & 3.2 < y_i^* \leq +\infty. \end{cases} \tag{4.1}$$

where threshold parameters are set as $\delta_1 = 0, \delta_2 = 1.6, \delta_3 = 3.2$. Quantile levels $K = 1, 3, 5, 7, 9$ are considered in following simulations.

4.2 Convergence diagnosis analysis

To guide the MCMC convergence, we first carry out a few test runs under the following settings of initial values for regression coefficients and prior hyperparameters using the Bayesian relative CQR approach for **Model 1** with $\varepsilon_i \sim N(0,1)$ and $K = 5$.

Setting 1: $\beta^{(0)} = (0,0,0,0,0,0)^T, a = b = 0.5, c = d = 0.5$.

Setting 2: $\beta^{(0)} = (1,1,1,1,1,1)^T, a = 1, b = 10, c = 10, d = 1$.

Setting 3: $\beta^{(0)} = (-1,-1,-1,-1,-1,-1)^T, a = 10, b = 1, c = 1, d = 10$.

For the above three settings, we run the Gibbs sampling algorithm 10000 iterations for each case. Figure 1 displays three MCMC chains of regression coefficients starting from the three settings in which the full mixing indicates a quick convergence of the MCMC algorithm. Figure 1 also shows that different initial values and priors do not produce a big impact on the algorithm convergence of Bayesian relative CQR approach. Hence, all following simulations are conducted based on the initial values and priors in **Setting 1**.

About the computation time, we conduct a test by taking the case of the error term t_3 and $K = 5$ as an example for two given models. The computing times of Bayesian relative CQR approach with $L_{1/2}$ penalty for accomplishing one replication by running 10000 times Gibbs sampling iterations are 3.97 minutes for **Model 1** and 5.61 minutes for **Model 2**. We see that Bayesian CQR approach with $K = 5$ for high-dimensional **Model 2** consumed more computational time than **Model 1**. Additionally, for the same model, the bigger for K, the more time the CQR approach will take. It should be noted that simulation studies in Section 4 and real-world data analysis in Section 5 are conducted using a Dell desktop [OptiPlex 7050, Intel(R) Core(TM) i7-7700U CPU] via statistical software R3.5.2. All codes of simulations and computations can be requested on the first author.

4.3 Substantive simulations

In this subsection, we conduct some simulations to compare estimation performance for the direct Bayesian CQR approach and Bayesian relative CQR. For **Model 1**, Figure 2

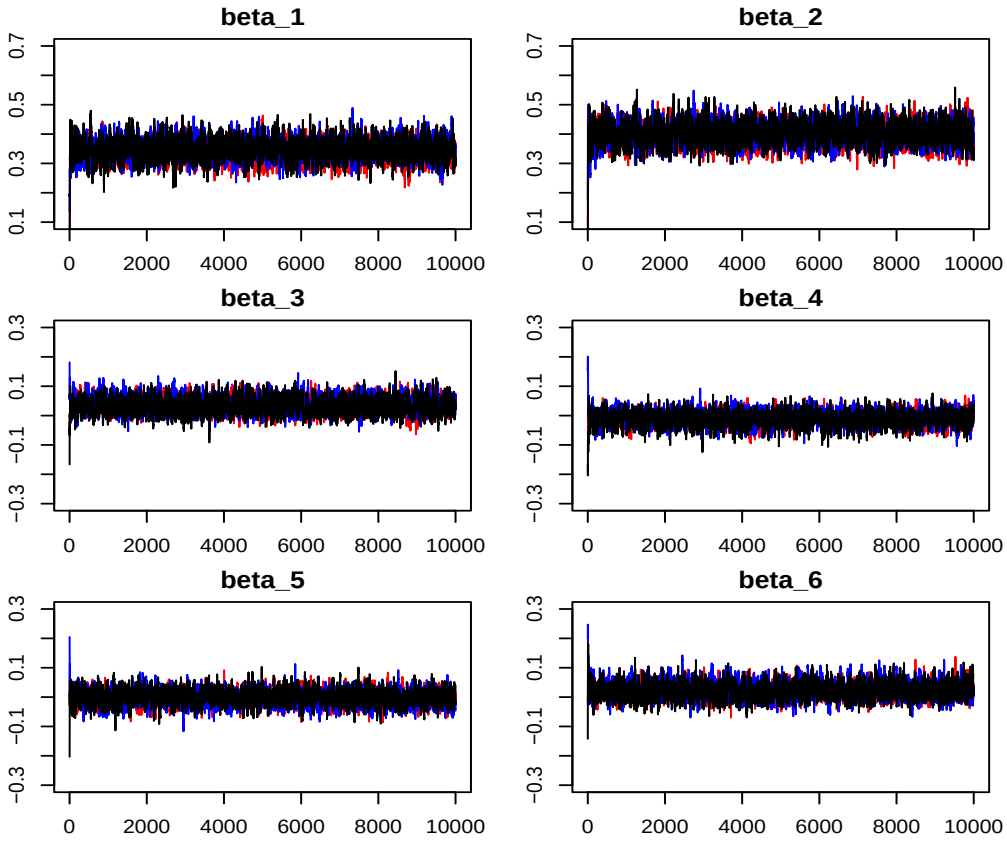

Figure 1: MCMC chains starting from three combinations of different initial values and prior hyperparameters for the Bayesian relative CQR. Notes: The red line denotes the MCMC plot of **Setting 1**; The blue line denotes the MCMC plot of the **Setting 2**; The black line denotes the **Setting 3**.

displays two MCMC chains of 10000 Gibbs samples of regression coefficients under the setting of t_3 error with $K = 5$, in which the red line denotes Bayesian CQR approach and the blue line denotes the Bayesian relative CQR approach. Figure 2 indicates that the MCMC algorithms of two class of Bayesian CQR approaches are convergent and Bayesian relative CQR approach has apparently better estimation performance. Similarly, for high-dimensional **Model 2**, Figure 3 indicates the Bayesian relative CQR approach has equally better estimation performance. To compare the estimation efficiency numerically between the direct Bayesian CQR approach and Bayesian relative CQR approach, Tables 1-2 listed the estimation bias (Bias), posterior root mean square error (RMSE), cover rate (CR) of 95% credible intervals, and variable selection results of 100 repeated simulations for two given models under two errors only considering $K = 5$. For each repetition, 10000 times Gibbs sampling algorithm is run for each combination, the previous 5000 burn-in samples are removed and the remaining 5000 samples are remained to conduct posterior inference. From Tables 1-2, we see that the Bayesian relative CQR approach has distinctly superior results to the direct Bayesian CQR approach. Additionally, Tables 1-2 also manifest that the Bayesian relative CQR approach produces robust results even for heavy-tailed t_3 error.

Next, we conduct simulation comparisons to illustrate the proposed Bayesian relative CQR over different K based on Bayesian $L_{1/2}$ ($\mathrm{BL}_{1/2}$) penalty and usual Bayesian LASSO (BLASSO) penalty. Based on 100 repeated simulations, the average posterior biases, and posterior RMSEs of relative regression parameters under the considered combinations are listed in Tables 3-4. Variable selection results of two Bayesian penalization approaches for **Models 1-2** are listed in Tables 5-6, where "NC" denotes the average correctly identified number of important covariates, and "NIC" denotes the average wrongly identified number of unimportant covariates. The correctly identified numbers of each parameter are also provided in Tables 5-6. The averaged posterior mean square error (APMSE) of the identified model for 100 simulations is given by

$$\mathrm{APMSE} = \frac{1}{100}\sum_{s=1}^{100}(\hat{\eta}^{(s)} - \eta)^T(\hat{\eta}^{(s)} - \eta), \tag{4.2}$$

where $\hat{\eta}^{(s)}$ is the $s-$th estimate of relative parameter η.

For **Model 1** and **Model 2**, from Tables 5-6, we see the Bayesian relative CQR

Table 1: Comparison for Bayesian relative CQR and Bayesian CQR results of **Model 1** based on $\mathrm{BL}_{1/2}$

K	Error	Methods	AMSE	NC	NIC	Evaluation	β_1	β_2	β_3	β_4	β_5	β_6	δ_3
5	normal	Relative CQR	0.006 (0.005)	2	0.00	bias	−0.004	0.002	−0.001	0.000	0.003	0.001	−1.108
						RMSE	0.048	0.045	0.020	0.022	0.021	0.019	1.147
						CR	0.90	0.90	1.00	0.98	0.98	0.99	0.06
						Variable selection	100	100	100	100	100	100	
		CQR	0.393 (0.118)	2	0.18	bias	−0.435	−0.421	−0.001	−0.001	0.006	0.003	−1.108
						RMSE	0.445	0.443	0.043	0.046	0.048	0.039	1.147
						CR	0.01	0.07	1.00	0.98	0.98	0.99	0.06
						Variable selection	100	100	98	95	92	97	
5	t_3	Relative CQR	0.010 (0.009)	2	0.07	bias	0.014	0.004	0.005	0.004	−0.003	−0.001	−1.695
						RMSE	0.059	0.053	0.030	0.036	0.029	0.029	1.705
						CR	0.86	0.87	0.97	0.93	0.98	0.96	0.00
						Variable selection	100	100	99	96	99	99	
		CQR	0.810 (0.154)	2	0.24	bias	−0.621	−0.636	0.007	0.006	−0.006	−0.002	−1.695
						RMSE	0.626	0.640	0.044	0.052	0.044	0.046	1.705
						CR	0.00	0.00	0.97	0.93	0.98	0.96	0.00
						Variable selection	100	100	95	91	95	95	

Table 2: Comparison of Bayesian relative CQR and Bayesian CQR results for **Model 2** with $\mathrm{BL}_{1/2}$

K	Error	Methods	AMSE	NC	NIC	Evaluation	β_1	β_2	β_3	β_5	β_{10}	β_{15}	β_{18}	β_{20}	δ_3
5	normal	Relative CQR	0.012 (0.009)	2	0.01	bias	−0.002	−0.010	0.003	0.002	0.000	0.001	0.002	0.001	−1.047
						RMSE	0.049	0.056	0.021	0.018	0.021	0.017	0.022	0.018	1.097
						CR	0.88	0.84	0.96	1.00	0.98	1.00	097	0.99	0.11
						Variable selection	100	100	100	100	100	100	100	100	
		CQR	0.399 (0.131)	2	0.75	bias	−0.409	−0.426	0.007	0.004	0.000	0.001	0.003	0.001	−1.047
						RMSE	0.421	0.439	0.043	0.039	0.048	0.035	0.049	0.038	1.097
						CR	0.04	0.05	0.96	1.00	0.98	1.00	0.97	0.99	0.11
						Variable selection	100	100	94	96	94	98	90	97	
5	t_3	Relative CQR	0.017 (0.010)	2	0.08	bias	−0.002	−0.001	−0.002	0.002	0.002	0.001	−0.002	0.000	−1.669
						RMSE	0.057	0.065	0.020	0.022	0.020	0.022	0.027	0.021	1.688
						CR	0.84	0.84	1.00	0.99	1.00	1.00	0.96	0.99	0.00
						Variable selection	100	100	100	100	100	100	99	100	
		CQR	0.850 (0.183)	2	0.53	bias	−0.637	−0.639	−0.003	0.004	0.002	0.002	−0.004	0.000	−1.669
						RMSE	0.644	0.644	0.030	0.034	0.030	0.034	0.041	0.033	1.688
						CR	0.00	0.00	1.00	0.99	1.00	1.00	0.96	0.99	0.00
						Variable selection	100	100	99	98	99	98	94	98	

Table 3: Estimates of Bayesian relative CQR for **Model 1** under different penalties

Methods	Error	K	Estimates	β_1	β_2	β_3	β_4	β_5	β_6	δ_3
$BL_{1/2}$	$N(0,1)$	$K=1$	Bias	−0.006	−0.005	−0.004	0.001	0.001	0.002	5.238
			RMSE	0.041	0.039	0.022	0.026	0.025	0.028	5.286
		$K=3$	Bias	−0.008	−0.004	0.001	−0.001	0.001	0.002	0.115
			RMSE	0.042	0.041	0.024	0.031	0.026	0.021	0.441
		$K=5$	Bias	−0.004	0.002	−0.001	0.000	0.003	0.001	−1.108
			RMSE	0.048	0.045	0.020	0.022	0.022	0.019	1.147
		$K=7$	Bias	−0.003	−0.006	0.001	0.001	−0.002	0.002	−1.597
			RMSE	0.065	0.062	0.024	0.021	0.025	0.022	1.623
		$K=9$	Bias	0.008	0.011	0.002	0.002	0.001	−0.005	−1.934
			RMSE	0.086	0.082	0.024	0.020	0.021	0.028	1.958
	t_3	$K=1$	Bias	−0.007	−0.006	0.002	0.002	−0.001	−0.009	2.996
			RMSE	0.047	0.045	0.025	0.024	0.026	0.030	3.067
		$K=3$	Bias	−0.006	−0.001	0.002	0.002	−0.002	0.000	−0.782
			RMSE	0.047	0.044	0.022	0.027	0.025	0.030	0.843
		$K=5$	Bias	0.014	0.004	0.005	0.004	−0.003	−0.001	−1.695
			RMSE	0.059	0.053	0.030	0.036	0.029	0.029	1.705
		$K=7$	Bias	0.008	0.000	0.001	−0.001	0.003	0.000	−2.088
			RMSE	0.077	0.079	0.028	0.026	0.033	0.031	2.099
		$K=9$	Bias	0.008	0.008	0.004	−0.006	−0.001	−0.003	−2.289
			RMSE	0.075	0.078	0.027	0.031	0.030	0.033	2.295
BLASSO	$N(0,1)$	$K=1$	Bias	−0.008	−0.005	0.000	−0.001	0.002	0.002	5.208
			RMSE	0.041	0.035	0.028	0.028	0.027	0.027	5.278
		$K=3$	Bias	−0.002	−0.010	−0.004	−0.002	0.000	0.000	−0.042
			RMSE	0.040	0.037	0.026	0.029	0.027	0.026	0.373
		$K=5$	Bias	−0.011	−0.004	0.004	−0.003	0.003	0.001	−1.041
			RMSE	0.050	0.049	0.024	0.026	0.027	0.030	1.082
		$K=7$	Bias	−0.011	−0.006	−0.002	0.006	0.003	−0.005	−1.596
			RMSE	0.057	0.062	0.025	0.031	0.027	0.025	1.622
		$K=9$	Bias	0.001	0.005	0.000	0.001	0.002	0.000	−1.935
			RMSE	0.072	0.070	0.030	0.028	0.029	0.026	1.951
	t_3	$K=1$	Bias	0.000	0.001	0.001	−0.004	0.000	0.007	2.974
			RMSE	0.041	0.044	0.032	0.036	0.031	0.038	3.043
		$K=3$	Bias	0.000	0.002	−0.004	−0.004	0.000	−0.001	−0.808
			RMSE	0.046	0.048	0.031	0.032	0.030	0.034	0.853
		$K=5$	Bias	−0.009	−0.009	0.003	0.001	−0.001	−0.002	−1.643
			RMSE	0.056	0.060	0.035	0.032	0.033	0.036	1.660
		$K=7$	Bias	−0.008	−0.003	−0.002	−0.002	−0.001	0.002	−2.052
			RMSE	0.075	0.069	0.038	0.034	0.033	0.034	2.061
		$K=9$	Bias	0.006	0.002	−0.001	−0.001	−0.007	−0.002	−2.271
			RMSE	0.088	0.090	0.035	0.038	0.034	0.036	2.279

Table 4: Estimates of Bayesian relative CQR for **Model 2** under different penalties

Method	Error	K	Est.	β_1	β_2	β_3	β_5	β_{10}	β_{15}	β_{18}	β_{20}	δ_3
$BL_{1/2}$	$N(0,1)$	$K=1$	Bias	−0.008	−0.014	−0.004	0.001	0.003	0.001	0.000	0.001	5.616
			RMSE	0.041	0.043	0.021	0.018	0.017	0.018	0.018	0.020	5.693
		$K=3$	Bias	−0.009	−0.008	−0.001	−0.002	0.004	0.001	−0.004	0.003	0.150
			RMSE	0.040	0.041	0.021	0.019	0.017	0.015	0.017	0.019	0.387
		$K=5$	Bias	−0.002	−0.010	0.003	0.002	0.000	0.001	0.002	0.001	−1.047
			RMSE	0.049	0.056	0.021	0.018	0.021	0.017	0.022	0.018	1.097
		$K=7$	Bias	−0.007	−0.006	0.001	−0.001	0.002	0.000	0.003	−0.001	−1.546
			RMSE	0.059	0.061	0.017	0.019	0.019	0.019	0.018	0.021	1.573
		$K=9$	Bias	0.016	0.009	0.003	−0.002	−0.001	0.002	−0.004	−0.001	−1.932
			RMSE	0.083	0.083	0.019	0.024	0.019	0.019	0.018	0.021	1.948
	t_3	$K=1$	Bias	−0.005	−0.011	0.001	0.001	0.002	−0.004	−0.003	−0.004	3.176
			RMSE	0.043	0.046	0.020	0.020	0.018	0.020	0.018	0.020	3.258
		$K=3$	Bias	−0.009	−0.005	−0.002	0.000	−0.001	−0.002	0.003	0.000	0.767
			RMSE	0.045	0.050	0.021	0.017	0.019	0.019	0.023	0.019	0.830
		$K=5$	Bias	−0.002	−0.001	−0.002	0.002	0.002	0.001	−0.002	0.000	−1.669
			RMSE	0.057	0.065	0.020	0.022	0.020	0.022	0.027	0.021	1.688
		$K=7$	Bias	0.018	−0.021	0.000	−0.001	0.001	0.000	−0.001	−0.001	−2.044
			RMSE	0.103	0.115	0.020	0.024	0.020	0.027	0.025	0.021	2.058
		$K=9$	Bias	0.009	−0.001	−0.002	0.001	−0.001	0.000	0.001	−0.004	−2.261
			RMSE	0.097	0.099	0.023	0.020	0.030	0.022	0.026	0.021	2.171
BLASSO	$N(0,1)$	$K=1$	Bias	−0.019	−0.019	0.001	0.000	−0.001	−0.001	0.000	0.001	5.495
			RMSE	0.045	0.048	0.24	0.026	0.024	0.025	0.025	0.022	5.521
		$K=3$	Bias	−0.013	−0.019	−0.001	0.001	−0.002	0.000	−0.002	−0.004	0.261
			RMSE	0.042	0.041	0.027	0.024	0.027	0.021	0.025	0.025	0.455
		$K=5$	Bias	−0.001	−0.008	0.001	0.003	0.000	−0.001	0.000	0.000	−1.006
			RMSE	0.048	0.057	0.026	0.026	0.026	0.027	0.025	0.022	1.058
		$K=7$	Bias	−0.007	−0.006	0.000	0.001	−0.001	−0.004	−0.001	−0.002	−1.555
			RMSE	0.057	0.063	0.026	0.024	0.028	0.022	0.027	0.023	1.579
		$K=9$	Bias	0.007	0.009	−0.003	0.000	0.000	−0.002	−0.001	0.000	−1.889
			RMSE	0.081	0.080	0.028	0.027	0.024	0.026	0.027	0.027	1.910
	t_3	$K=1$	Bias	−0.011	−0.020	0.003	−0.003	0.003	0.003	0.002	0.002	3.185
			RMSE	0.054	0.051	0.030	0.026	0.024	0.030	0.028	0.029	3.261
		$K=3$	Bias	−0.013	−0.011	0.002	−0.004	−0.004	−0.001	0.001	−0.002	−0.771
			RMSE	0.045	0.049	0.031	0.030	0.025	0.031	0.027	0.028	0.808
		$K=5$	Bias	−0.013	−0.006	0.000	−0.003	−0.002	0.002	−0.001	−0.001	−1.591
			RMSE	0.050	0.055	0.028	0.029	0.030	0.034	0.029	0.028	1.606
		$K=7$	Bias	0.004	−0.006	−0.003	−0.001	0.004	−0.001	−0.003	0.000	−2.031
			RMSE	0.068	0.070	0.033	0.032	0.034	0.032	0.031	0.031	2.041
		$K=9$	Bias	0.006	0.011	−0.010	0.002	0.003	0.002	0.000	−0.002	−2.271
			RMSE	0.091	0.080	0.034	0.031	0.033	0.029	0.030	0.032	2.278

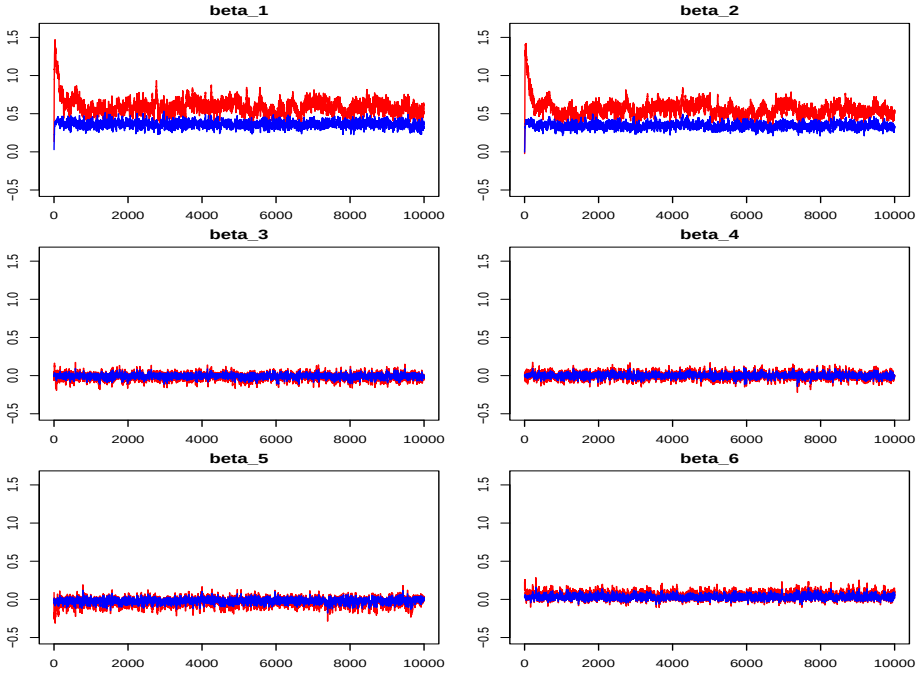

Figure 2: MCMC trace plots of 10000 Gibbs samples for **Model 1** under t_3 error for K=5 based on $\text{BL}_{1/2}$ estimation approach. Notes: The red line denotes direct Bayesian CQR estimates and the blue line denotes Bayesian relative CQR estimates.

Table 5: Variable selections of Bayesian relative CQR for **Model 1** under two penalties

Methods	Error	K	APMSE	NC	NIC	β_1	β_2	β_3	β_4	β_5	β_6
$\text{BL}_{1/2}$	$N(0,1)$	$K=1$	0.006 (0.004)	2	0.02	100	100	100	100	100	98
		$K=3$	0.006 (0.005)	2	0.01	100	100	100	99	100	100
		$K=5$	0.006 (0.005)	2	0.01	100	100	100	100	100	100
		$K=7$	0.010 (0.016)	2	0	100	100	100	100	100	100
		$K=9$	0.016 (0.027)	2	0.01	100	100	100	100	100	99
	t_3	$K=1$	0.007 (0.005)	2	0.01	100	100	99	100	100	100
		$K=3$	0.007 (0.005)	2	0.03	100	100	100	99	99	99
		$K=5$	0.010 (0.009)	2	0.07	100	100	99	96	99	99
		$K=7$	0.015 (0.017)	2	0.02	100	100	99	100	99	100
		$K=9$	0.015 (0.018)	2	0.04	100	100	99	98	100	99
BLASSO	$N(0,1)$	$K=1$	0.006 (0.004)	2	0.02	100	100	100	99	99	100
		$K=3$	0.006 (0.004)	2	0.00	100	100	100	100	100	100
		$K=5$	0.008 (0.006)	2	0.01	100	100	100	100	100	99
		$K=7$	0.010 (0.010)	2	0.00	100	100	100	100	100	100
		$K=9$	0.013 (0.011)	2	0.00	100	100	100	100	100	100
	t_3	$K=1$	0.008 (0.005)	2	0.05	100	100	100	99	100	96
		$K=3$	0.008 (0.005)	2	0.03	100	100	100	100	98	99
		$K=5$	0.011 (0.009)	2	0.05	100	100	98	98	100	99
		$K=7$	0.015 (0.014)	2	0.05	100	100	99	98	99	99
		$K=9$	0.021 (0.025)	2	0.04	100	100	99	97	100	100

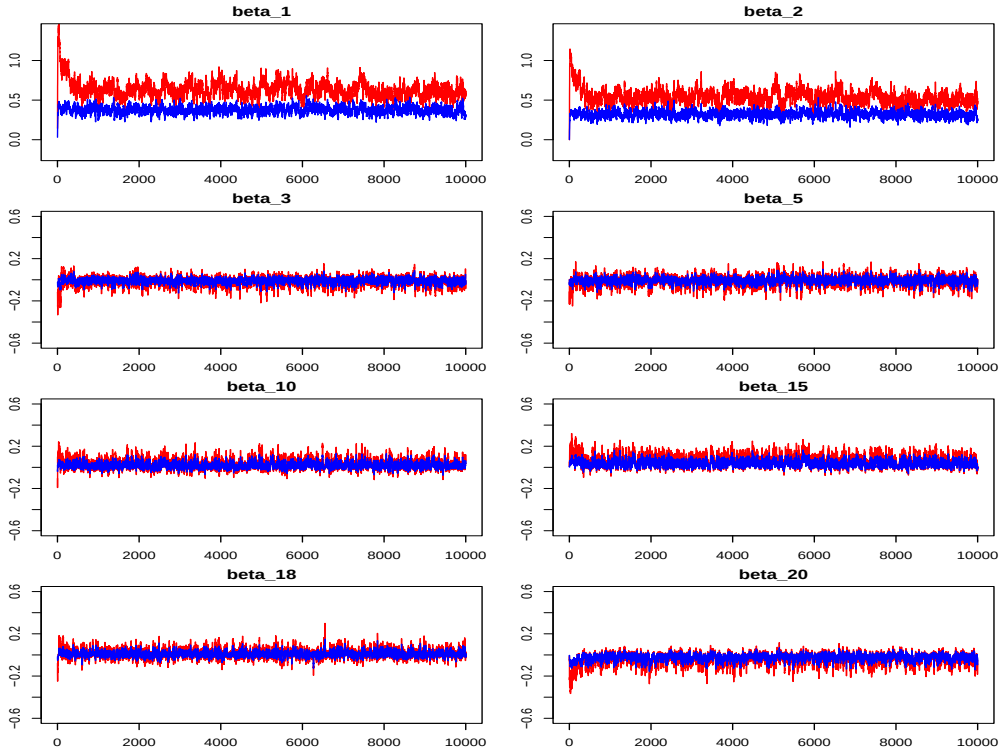

Figure 3: MCMC trace plots of 10000 Gibbs samples for **Model 2** under t_3 error for K=5 based on $\text{BL}_{1/2}$ estimation approach. Notes: The red line denotes direct Bayesian estimates and the blue line denotes Bayesian relative CQR estimates.

Table 6: Variable selections of Bayesian relative CQR for **Model 2** under two penalties

Methods	Error	K	APMSE(β/δ_3)	NC	NIC	β_1	β_2	β_3	β_5	β_{10}	β_{15}	β_{18}	β_{20}
$BL_{1/2}$	$N(0,1)$	$K=1$	0.009 (0.005)	2	0.02	100	100	100	100	100	100	99	99
		$K=3$	0.009 (0.005)	2	0.02	100	100	100	100	100	100	100	100
		$K=5$	0.012 (0.009)	2	0.01	100	100	100	100	100	100	100	100
		$K=7$	0.013 (0.011)	2	0.02	100	100	100	99	100	100	100	100
		$K=9$	0.020 (0.032)	2	0.04	100	100	100	100	100	100	100	100
	t_3	$K=1$	0.011 (0.006)	2	0.02	100	100	100	100	100	100	100	99
		$K=3$	0.012 (0.006)	2	0.04	100	100	100	100	100	100	99	100
		$K=5$	0.017 (0.010)	2	0.08	100	100	100	100	100	100	99	100
		$K=7$	0.033 (0.096)	2	0.07	100	100	100	99	99	99	99	100
		$K=9$	0.029 (0.042)	2	0.12	100	100	100	100	97	100	99	100
BLASSO	$N(0,1)$	$K=1$	0.014 (0.006)	2	0.01	100	100	100	100	100	100	99	100
		$K=3$	0.014 (0.005)	2	0.05	100	100	100	100	99	100	100	100
		$K=5$	0.017 (0.009)	2	0.06	100	100	99	99	99	100	100	100
		$K=7$	0.018 (0.012)	2	0.03	100	100	99	100	99	100	100	100
		$K=9$	0.026 (0.018)	2	0.09	100	100	99	100	100	100	98	100
	t_3	$K=1$	0.020 (0.009)	2	0.09	100	100	100	100	100	99	100	99
		$K=3$	0.019 (0.007)	2	0.06	100	100	98	99	100	100	100	100
		$K=5$	0.021 (0.010)	2	0.08	100	100	100	99	99	99	100	100
		$K=7$	0.027 (0.016)	2	0.18	100	100	98	99	99	99	99	98
		$K=9$	0.033 (0.030)	2	0.16	100	100	97	98	100	99	100	99

approach with $L_{1/2}$ penalty can specify all parameters more accurately with the same "NC" but smaller "NIC" than Bayesian relative CQR with LASSO penalty under two error distributions over different K. We conclude that the Bayesian $L_{1/2}$ relative CQR approach has better estimation performance than the Bayesian LASSO relative CQR approach. Hence, the Bayesian $L_{1/2}$ relative CQR is consistently recommended due to its good performance. Additionally, for different K, the Bayesian $L_{1/2}$ relative CQR approach has different performance. From the estimation results in Tables 3-4, we see that although regression coefficients can be precisely estimated for almost all of K, the Bias and RMSE of threshold parameter δ_3 are the smallest consistently only for $K = 3$. Comprehensively, the Bayesian $L_{1/2}$ relative CQR with $K = 3$ is recommended for all the considered models.

5 Real data example

We analyze a knowledge level dataset in Kahraman et al. (2013) in this section. It is the real dataset about the students' knowledge status about electrical DC machines. The

dataset includes 403 instances and five attribute variables. The users' knowledge class was classified into four levels by the authors using an intuitive knowledge classifier: very low (50 persons), low (129 persons), middle (122 persons), and high (130 persons). The description of variables for the knowledge level is listed in Table 7. We analyze this dataset using the recommended Bayesian $L_{1/2}$ relative CQR approach with $K = 3$ to explore which attributes have a significant impact on ordinal classification outcomes. The initial values and priors of parameters are set the same to the simulations in Subsection 4.3. We run 10000 times Gibbs sampling algorithm to conduct Bayesian posterior inference, the previous 5000 burn-in samples are removed and the remaining 5000 samples are used to calculate posterior estimates (Est.), standard deviation (St.d), 95% credible lower bound (LB) and upper bound (UB). All computational results based on direct Bayesian CQR and Bayesian relative CQR are listed in Table 8. The results of **Restored** CQR estimates of original parameters based on Bayesian relative CQR approach are also listed in Table 8 for the aim of comparison. Table 8 shows that the Bayesian relative CQR approach produces better estimation results with smaller St.ds and shorter 95% interval lengths than the direct Bayesian CQR approach. Additionally, although there is no overt difference between the estimation values of the **Restored** approach based on Bayesian relative CQR and direct Bayesian CQR approach, the former CQR approach has smaller St.ds and shorter 95% interval lengths for all parameters which brings about more robust results.

The attribute **STG** have a negative effect on the knowledge level of use, while **SCG**, **STR**, **LPR** and **PEG** have positive effects on response UNS. Additionally, according to the 95% Bayesian credible intervals, we conclude that **SCG**, **STR**, and **PEG** are significant attributes on the user's knowledge level.

6 Conclusion

This paper propose a Bayesian $L_{1/2}$ relative CQR approach for the latent ordinal regression model. Monte Carlo simulations and a real data example are implemented to illustrate the proposed procedures. Based on all simulations and real data analysis, we conclude that Bayesian $L_{1/2}$ relative CQR approach with $K = 3$ can accurately specify important predictors for latent ordinal regression models. The suggested approach contribute robust

Table 7: Description of variables for the knowledge level data

Variable	Definition	Description
Response	UNS	The knowledge level of user, 1=Very Low, 2=Low, 3=Middle, 4=high.
Covariates	STG	The degree of study time for goal object materails
	SCG	The degree of repetition number of user for goal object materails
	STR	The degree of study time of user for related objects with goal object
	LPR	The exam performance of user for related objects with goal object
	PEG	The exam performance of user for goal objects

Table 8: Estimates of Bayesian relative CQR and direct CQR of knowledge level data

Method	Estimation	**STG**	**SCG**	**STR**	**LPR**	**PEG**	δ_1	δ_2	δ_3
CQR	Est.	−0.259	1.564	0.865	6.254	19.97	6.385	10.87	16.65
	St.d	0.555	0.628	0.516	0.798	1.757	0.869	1.188	1.645
	95%LB	−1.319	0.064	−0.259	4.568	15.350	4.323	8.059	12.546
	95%UB	0.918	2.625	1.725	7.653	22.365	7.710	12.615	19.116
Relative CQR	Est.	−0.017	0.093	0.051	0.375	1.201	−−	−−	−−
	St.d	0.034	0.035	0.030	0.029	0.045	−−	−−	−−
	95%LB	−0.088	0.004	−0.016	0.319	1.122	−−	−−	−−
	95%UB	0.052	0.147	0.099	0.432	1.302	−−	−−	−−
	Restored Est.	−0.283	1.553	0.852	6.251	20.00			

estimation results even for non-normal latent regression models and can be naturally extended to ordinal longitudinal data models.

Acknowledgements

Authors thank editors and referees for their constructive comments and suggestions which have greatly improved the paper. The research was supported by grants from the National Natural Science Foundation of China (grant 12061065) and Funds for Innovative Fundamental Research Group Project of Gansu Province of China (grant 23JRRA684).

References

Agresti, A (2010). Analysis of ordinal categorical data (2nd ed.). New York: Wiley.

Alhamzawi, R (2016). Bayesian analysis of composite quantile regression. Statistics in Biosciences, 8(2), 1-16.

Alhamzawi, R., Ali, H. T. M (2018). Bayesian Tobit quantile regression with $L_{1/2}$ penalty. Communications in Statistics-Simulation and Computation, 47(6), 1739-1750.

Alhamzawi, R., Yu, K., Benoit, D. F (2012). Bayesian adaptive Lasso quantile regression. Statistical Modelling, 12(3), 279-297.

Cliff, N (1996). Ordinal methods for behavioral data analysis. Lawrence Erlbaum Associates.

Davino, C., Furno, M., Vistocco, D (2014). Quantile regression: theory and applications. New York: John Wiley & Sons.

Fan, J., Li, R (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96(456), 1348-1360.

Frank, A., Asuncion, A (2010). UCI machine learning repository. University of California, Irvine.

Fu, W. J (1998). Penalized regression: the bridge versus lasso. Journal of Computational and Graphical Statistics, 7(3), 397-416.

Grabski, I. N., Vito, R. D., Engelhardt, B. E (2019). Bayesian ordinal quantile regression with a partially collapsed gibbs sampler. doi.10.48550/arXiv.1911.07099.

Huang, H., Chen, Z (2015). Bayesian composite quantile regression. Journal of Statistical Computation & Simulation, 85(18), 1-11.

Huang, X., Zhan, Z (2022). Local composite quantile regression for regression discontinuity. Journal of Business & Economic Statistics, 40(4), 1863-1875.

Hui, Z., Hastie, T (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, 67(5), 768-768.

Jiang, J. C., Jiang, X. J., Song, X. Y (2014). Weighted composite quantile regression estimation of DTARCH models. The Econometrics Journal, 17(1), 1-23.

Kai, B., Li, R. Z., Zou, H (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. Journal of the Royal Statistical Society: Series B, 72, 49-69.

Kahraman, H. T., Sagiroglu, S., Colak, I (2013). Developing intuitive knowledge classifier and modeling of users' domain dependent data in web. Knowledge Based Systems, 37, 283-295.

Manuguerra, M., Heller, G. Z (2010). Ordinal regression models for continuous scales. The International Journal of Biostatistics, 6(1), Article 14.

Mallick, H., Yi, N (2018). Bayesian bridge regression. Journal of Applied Statistics, 45(6), 988-1008.

Montesinos-Lopez, O. A., Montesinos-Lopez, A., Crossa, J., et al. (2015). Genomic-enabled prediction of ordinal data with Bayesian logistic ordinal regression. Genomic Selection, 5(10), 2113-2126.

Park, T., Casella, G (2008). The Bayesian Lasso. Journal of the American Statistical Association, 103(482), 681-686.

Polson, N. G., Scott, J. G., Windle, J (2014). The Bayesian bridge. Journal of the Royal Statistical Society: Series B, 76(4), 713-733.

Sha, N., Dechi, B. O (2019). A bayes inference for ordinal response with latent variable approach. Statistics, 2(2), 321-331.

Tang, L. J., Zhou, Z. G. and Wu, C. C (2012). Weighted composite quantile estimation and variable selection method for censored regression model. Statistics and Probability Letters, 82, 653-663.

Tian, Y. Z., Lian, H., Tian, M. Z (2017). Bayesian composite quantile regression for linear mixed effects models. Communications in Statistics-Theory and Method, 15(46), 7717-7731.

Tian, Y. Z., Wang, L. Y., Tang, M. L., Tian, M. Z (2021). Weighted composite quantile regression for longitudinal mixed effects models with application to AIDS studies. Communications in Statistics-Simulation and Computation, 50(6), 1837-1853.

Tian, Y. Z., Zhu, Q. Q., Tian, M. Z (2016). Estimation of linear composite quantile regression using EM algorithm. Statistics & Probability Letters, 117, 183-191.

Tian, Y. Z., Wang, L. Y., Wu, X. Q., Tian, M. Z (2019). Gibbs sampler algorithm of Bayesian weighted composite quantile regression. Chinese Journal of Applied Probability and Statistics, 35(2), 178-192.

Tibshirani, R (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B, 58(1), 267-288.

Tutz, G (2022). Ordinal regression: a review and a taxonomy of models. WIREs Computational Statistics, 14(2), e1545.

Wang, M., Chen, Z., Wang, C. D (2018). Composite quantile regression for GARCH models using high-frequency data. Econometrics and Statistics, 7, 115-133.

Xu, Z., Zhang, H., Wang, Y., Chang, X., Liang, Y (2010). $L_{1/2}$ regularization. Science China Information Sciences, 53(6), 1159-1169.

Zhang, H. P., Feng, R., Zhu, H. T (2003). A latent variable model of segregation analysis for ordinal traits. Journal of the American Statistical Association, 98(464), 1023-1034.

Zou, H (2006). The adaptive LASSO and its oracle properties. Journal of the American Statistical Association, 101(476), 1418-1429.

Zou, H., Yuan, M (2008). Composite quantile regression and the oracle model selection theory. The Annals of Statistics, 36, 1108-1126.