Statistics and Its Interface Volume 13 (2020) 271–286

Fully Bayesian $L_{1/2}$-penalized linear quantile regression analysis with autoregressive errors

Yuzhu Tian and Xinyuan Song*

In the quantile regression framework, we incorporate Bayesian $L_{1/2}$ and adaptive $L_{1/2}$ penalties into quantile linear regression models with autoregressive (AR) errors to conduct statistical inference. A Bayesian joint hierarchical model is established using the working likelihood of the asymmetric Laplace distribution (ALD). On the basis of the mixture representations of ALD and the generalized Gaussian distribution priors of regression coefficients and AR parameters, a Markov chain Monte Carlo algorithm is developed to conduct posterior inference. Finally, the proposed Bayesian estimation procedures are demonstrated by simulation studies and applied to a real data application concerning the electricity consumption of residential customers.

AMS 2000 subject classifications: 62F15, 62J07.
Keywords and phrases: Autoregressive error, Bayesian quantile regression, Generalized Gaussian distribution, Gibbs sampler, $L_{1/2}$ penalty.

1. INTRODUCTION

Early developments of regression analysis frequently assume that the error terms are independent. However, this independence assumption is likely to be violated in substantive research. Therefore, regression with dependent errors has attracted considerable attention in the statistical literature. For example, Tasy (1984) and Lin et al. (2000) proposed linear regression models with time series errors. Lee and Lund (2004) discussed a linear regression with stationary autocorrelated errors. Yang (2012) incorporated serially correlated errors into regression. Rosadi and Peiris (2015) investigated a second-order least-squares regression with autocorrelated errors. Nevertheless, a majority of classical regression analyses have focused on mean regression, which considers only the average of the response variable conditional on covariates and is expected to be sensitive to outliers or non-normal errors. Koenker and Bassett (1978) proposed quantile regression (QR) as a comprehensive alternative to mean regression. In the past decades, QR has become an appealing statistical modeling tool in modern regression analysis because it allows to depict the effect of covariates on the whole conditional distributions of the response variable instead of only on its average. A comprehensive overview of QR can be found in Koenker (2005) and Davino et al. (2014). Meanwhile, Yu and Moyeed (2001) introduced QR techniques into the Bayesian framework. Later on, Geraci and Bottai (2007) conduct Bayesian QR analysis for longitudinal data using the asymmetric Laplace distribution (ALD). Reich et al. (2011) proposed Bayesian spatial QR models. Zhao and Lian (2015) investigated Bayesian Tobit QR with single-index models. Wang et al. (2015) considered Bayesian quantile structural equation models. Tian et al. (2016) proposed Bayesian joint QR for mixed effects models with censoring and errors in covariates. Huang and Chen (2016) studied Bayesian QR-based nonlinear mixed-effects joint models for time-to-event and longitudinal data with multiple features. Despite the fruitful literature in QR analysis, the preceding studies concentrated on estimation.

Variable selection is important in obtaining parsimonious models that retain only significant covariates. A variety of variable selection methods have been developed in the literature. These methods include classical criterion-based procedures, such as Akaike information criterion (AIC) and Bayesian information criterion (BIC), and modern regularization methods, such as ridge penalty (Hoerl and Kennard, 1970), least absolute shrinkage and select operator (LASSO; Tibshirani, 1996), adaptive LASSO (Zou, 2006), smoothly clipped absolute deviation (SCAD; Fan and Li, 2001), elastic-net penalty (Enet; Zou and Hastie, 2005), bridge penalty (Fu, 1998; Knight and Fu, 2000), and $L_{1/2}$-norm penalty (Xu et al., 2010). As a representative penalization approach, an L_ξ-norm regularized penalty includes the best subset selection ($\xi = 0$), LASSO regression ($\xi = 1$), ridge regression ($\xi = 2$), bridge regression ($0 < \xi < 1$), and $L_{1/2}$ regularizer ($\xi = 1/2$) as special cases. Among the L_ξ-norm regularizers, bridge regression ($0 < \xi < 1$), especially $L_{1/2}$, has many desirable statistical properties. A theoretical justification provided by Xu et al. (2010) shows that $L_{1/2}$ is the most sparse and robust among the L_ξ ($0 < \xi \le 1$) regularizers. Given such nice properties, the $L_{1/2}$-based regularized methods have been applied to various statistical models. For example, Liang et al. (2013) studied sparse logistic regression with an $L_{1/2}$ penalty. Zhang et al. (2014) considered a sharp nonasymptotic bound and phase diagram of $L_{1/2}$ regularization. Luan et al. (2014) developed the $L_{1/2}$ regularization shooting method for Cox's proportional hazards model. However, these developments were restricted

*Corresponding author.

within mean regression with independent errors in a frequentist framework. No existing study has ever investigated the $L_{1/2}$ regularization method for QR with dependent errors in a Bayesian framework.

In this study, we aim to fill this gap and consider a $L_{1/2}$ penalized linear QR model with autoregressive (AR) errors. We propose a fully Bayesian approach because of its potential of managing highly complex data/model structure and the capability of incorporating additional model inputs that stem from the prior distributions of model parameters. Various regularized penalization methods have been available either for mean regression models with dependent errors (Wang et al., 2007; Wu and Wang, 2012; Fallahpour and Ahmed, 2014; Yoon et al., 2017) or for Bayesian regression models with independent errors (Park and Casella, 2008; Yu et al., 2013; Polsonet al., 2014; Alhamzawi and Ali, 2018; Mallick and Yi, 2018; Kang et al., 2019). However, the available methods cannot simultaneously accommodate all the features of the proposed model. To our knowledge, we are the first to introduce the $L_{1/2}$ penalization to the Bayesian QR model with AR errors.

Several recent works also considered regularized methods for linear QR models with AR errors (Lim and Oh, 2014; Jiang and Li, 2014; Tian et al., 2019). Despite a similarity in the model framework, the present study differs from the previous works in three aspects. First, the penalties considered are different. Unlike the previous studies that focused on SCAD (Lim and Oh, 2014) or LASSO penalties (Jiang and Li, 2014; Tian et al., 2019), this study considers the $L_{1/2}$ penalty, which has been demonstrated to be the most sparse and robust among L_ξ $(0 < \xi \leq 1)$ in the context of linear mean regression models (Xu et al., 2010). We are interested to investigate whether such attractive properties of the $L_{1/2}$ regularizer retain for Bayesian QR analysis in the presence of dependent errors. Second, the inference procedures are different. Lim and Oh (2014) and Jiang and Li (2014) developed SCAD- or LASSO-based penalized procedures from a frequentist perspective. Although Tian et al. (2019) considered a Bayesian approach, they focused also on LASSO-type penalties. Given different penalized methods considered, the proposed posterior inferences including the prior specification, posterior derivation, and the posterior sampling are different from those of Tian et al. (2019). In particular, we utilize the mixture uniform-Gamma representation of the generalized Gaussian distribution (GDD) proposed by Mallick and Yi (2018) to simplify the prior specification and posterior derivation of the regression coefficients, thereby facilitating a simple and efficient computing algorithm. Our simulation studies (Section 5) demonstrate that the proposed $L_{1/2}$-based penalized methods are faster than the LASSO-based methods of Tian et al. (2019). Finally, the real application results are different. In comparison with Lim and Oh (2014) and Tian et al. (2019) that analyzed exactly the same dataset, our study provides new insights into the quantile-specific influential factors of the response variable (Section 6).

The remainder of this paper is organized as follows. Section 2 presents the hierarchical model and the joint likelihood function. Section 3 introduces the Bayesian $L_{1/2}$ penalized QR (B$L_{1/2}$) of the proposed model. Section 4 presents an adaptive version of B$L_{1/2}$ (BA$L_{1/2}$). Section 5 conducts simulation studies to demonstrate the empirical performance of the proposed methods. Section 6 presents the application of the proposed methodology to a real-life study. Section 7 concludes the paper.

2. THE MODEL AND WORKING LIKELIHOOD

We consider a linear regression model with AR(q) error as follows:

$$(1)\quad y_t = x_t^T\beta + \varepsilon_t, \varepsilon_t = \sum_{j=1}^{q} \phi_j \varepsilon_{t-j} + \eta_t,\ t = q+1, \cdots, n,$$

where $x_t = (x_{t1}, \cdots, x_{tp})^T$ is a p-dimensional covariate, $\beta = (\beta_1, \cdots, \beta_p)^T$ is the regression coefficient, p is the dimension of x_t, q is the AR order, which is determined either through model selection criterion, such as AIC and BIC, or based on the autocorrelation function (ACF) and partial ACF (PACF) plots of the model residuals, $\Phi = (\phi_1, \cdots, \phi_q)^T$ is the AR coefficient, η_t is an independently and identically distributed (i.i.d.) error term. The condition of stationarity for AR(q) error ε_t in model (1) is that all inverse characteristic roots of the polynomial $1 - \sum_{j=1}^{q} \phi_j z^j$ are inside the unit circle.

For any given quantile level $\tau \in (0, 1)$, according to Koenker and Bassett (1978), the τth QR estimators $\hat{\beta}$ and $\hat{\Phi}$ of model (1) can proceed by the solution to minimize the following objective loss function:

$$(2)\ Q_0(\beta, \Phi) = \sum_{t=q+1}^{n} \rho_\tau[(y_t - x_t^T\beta) - \sum_{j=1}^{q} \phi_j (y_{t-j} - x_{t-j}^T\beta)],$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile check function. Notably, the loss function (2) is non-convex, non-differentiable, and multimodal. Thus, directly minimizing the loss function to perform parameter estimation is infeasible.

Yu and Moyeed (2001) suggested a Bayesian approach to conduct QR estimation by assuming ALD for the model error as a working model. On the basis of such an assumption, we know that maximizing the working likelihood under ALD errors of model (1) is equivalent to minimizing the QR objective loss function (2). The probability density function of ALD for the error term η_t is written explicitly as

$$(3)\qquad f(\eta_t|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\Big\{ -\rho_\tau(\frac{\eta_t - \mu}{\sigma})\Big\},$$

where μ is the location, and σ is the auxiliary scale parameter.

Then, the conditional working likelihood of QR model (2) can be equivalently written as follows:

$$\text{(4)}\quad \begin{aligned}&\mathrm{H}(\beta,\Phi,\sigma|(x_1^T,y_1),\cdots,(x_n^T,y_n))=\prod_{t=q+1}^{n}\frac{\tau(1-\tau)}{\sigma}\\&\cdot\exp\Big\{-\rho_\tau\Big[\frac{y_t-x_t^T\beta-\sum_{j=1}^{q}\phi_j(y_{t-j}-x_{t-j}^T\beta)}{\sigma}\Big]\Big\}.\end{aligned}$$

However, we cannot derive tractable estimators based on the aforementioned working likelihood because of computational difficulty with the inherent non-differentiability. Fortunately, we can utilize a Gaussian mixture representation of ALD proposed by Reed and Yu (2009) and Kozumi and Kobayashi (2011) to overcome this intractability. Using the mixture representation, we can represent model (1) as

$$\text{(5)}\quad \begin{aligned}y_t=x_t^T\beta+\sum_{j=1}^{q}\phi_j(y_{t-j}-x_{t-j}^T\beta)+\theta_1\upsilon_t+\sqrt{\theta_2\sigma\upsilon_t}e_t,\\ t=q+1,\cdots,n,\end{aligned}$$

where $\theta_1=\frac{1-2\tau}{\tau(1-\tau)},\theta_2=\frac{2}{\tau(1-\tau)},\upsilon_t\sim\mathrm{Exp}(\frac{1}{\sigma})$, $\mathrm{Exp}(\cdot)$ denotes the exponential distribution, $e_t\sim N(0,1)$, and υ_t and e_t are independent. From (5), the conditional distribution of y_t is normal with mean $x_t^T\beta+\sum_{j=1}^{q}\phi_j(y_{t-j}-x_{t-j}^T\beta)+\theta_1\upsilon_t$ and variance $\theta_2\sigma\upsilon_t$.

Let $\upsilon=(\upsilon_1,\cdots,\upsilon_n)$. Based on the mixture representation (5), we can express the joint working likelihood of the data $\{y,x,\upsilon\}$ as

(6)
$$\begin{aligned}&\mathrm{L}(\beta,\Phi,\sigma|y,x,\upsilon)=\prod_{t=q+1}^{n}\Big[\frac{1}{\sqrt{2\pi}\cdot\sqrt{\theta_2\sigma\upsilon_t}}\\&\cdot\exp\Big\{-\frac{(y_t-x_t^T\beta-\sum_{j=1}^{q}\phi_j(y_{t-j}-x_{t-j}^T\beta)-\theta_1\upsilon_t)^2}{2\theta_2\sigma\upsilon_t}\Big\}\\&\cdot\frac{1}{\sigma}\exp\Big\{-\frac{1}{\sigma}\upsilon_t\Big\}\Big].\end{aligned}$$

For QR model (2), to select significant regression coefficients and AR parameters, we can employ regularized estimation with proper penalty functions. The penalized objective loss function of model (2) can be given as follows:

$$\text{(7)}\quad Q_1(\beta,\Phi)=Q_0(\beta,\Phi)+\sum_{i=1}^{p}p_\lambda(|\beta|_i)+\sum_{j=1}^{q}p_\gamma(|\varphi|_j),$$

where $\lambda>0$ and $\gamma>0$ are penalty indexes. The $L_{1/2}$ penalty is imposed on the regression coefficients and AR parameters as follows:

$$p_\lambda(|\beta|_i)=\lambda|\beta_i|^{1/2},\ p_\gamma(|\phi_j|)=\gamma|\phi_j|^{1/2}.$$

3. BAYESIAN $L_{1/2}$ PENALIZED QR (B$L_{1/2}$)

3.1 Prior distributions

To conduct the Bayesian analysis of the $L_{1/2}$ penalized QR model (7), we specify GGD priors for the regression coefficients and AR parameters as follows:

$$\begin{aligned}\pi(\beta|\sigma,\lambda)&=\prod_{i=1}^{p}\pi(\beta_i|\sigma,\lambda),\\ \pi(\beta_i|\sigma,\lambda)&=\frac{\lambda^2}{4\sigma^2}\exp\Big\{-\frac{\lambda}{\sigma}|\beta_i|^{1/2}\Big\};\\ \pi(\Phi|\sigma,\gamma)&=\prod_{j=1}^{q}\pi(\phi_j|\sigma,\gamma),\\ \pi(\phi_j|\sigma,\gamma)&=\frac{\gamma^2}{4\sigma^2}\exp\Big\{-\frac{\gamma}{\sigma}|\phi_j|^{1/2}\Big\};\end{aligned}$$

where $\lambda>0$ and $\gamma>0$ are defined in (7).

Incorporating the GGD priors into the working likelihood function (4) results in a marginal posterior density as follows:

(8)
$$\begin{aligned}&\pi(\beta,\Phi|(x_1^T,y_1),\cdots,(x_n^T,y_n))\\&=\exp\Big\{-\sum_{t=q+1}^{n}\rho_\tau\Big[\frac{y_t-x_t^T\beta-\sum_{j=1}^{q}\phi_j(y_{t-j}-x_{t-j}^T\beta)}{\sigma}\Big]\Big\}\\&\quad\cdot\exp\Big\{-\frac{\lambda}{\sigma}\sum_{i=1}^{p}|\beta_i|^{1/2}-\frac{\gamma}{\sigma}\sum_{j=1}^{q}|\phi_j|^{1/2}\Big\}\\&=\exp\Big\{-\frac{1}{\sigma}Q_1(\beta,\Phi)\Big\}.\end{aligned}$$

The penalized estimators of β and Φ from (7) amount to the posterior modes of the marginal posterior density (8). However, the posterior distribution in (8) is analytically intractable. Polson et al. (2014) proposed a mixture representation of GGD to conduct Bayesian bridge regression. Mallick and Yi (2018) further suggested a simple and efficient mixture uniform-Gamma representation of GGD as follows:

$$\begin{aligned}&\frac{\lambda^{1/\alpha}}{2\Gamma(1+1/\alpha)}\exp\big\{-\lambda|x|^\alpha\big\}\\&=\int_{|x|^\alpha}^{\infty}\frac{1}{2u^{1/\alpha}}\frac{\lambda^{1+1/\alpha}}{\Gamma(1+1/\alpha)}u^{1/\alpha}\exp\big\{-\lambda u\big\}du.\end{aligned}$$

By using the above result and setting $\alpha=1/2$, we can decompose the priors of β_i and ϕ_j as follows:

$$\begin{aligned}\pi(\beta_i|\sigma,\lambda)&=\int_0^\infty\pi(\beta_i|\lambda,\sigma,s_i)\pi(s_i)ds_i,\\ \pi(\phi_j|\sigma,\gamma)&=\int_0^\infty\pi(\phi_j|\gamma,\sigma,\omega_j)\pi(\omega_j)d\omega_j,\end{aligned}$$

where

$$\pi(\beta_i|\lambda,\sigma,s_i)=\text{Uniform}(-s_i^2,s_i^2),$$
$$\pi(s_i)=\text{Gamma}\Big(3,\frac{\lambda}{\sigma}\Big), i=1,\cdots,p;$$
$$\pi(\phi_j|\gamma,\sigma,\omega_j)=\text{Uniform}(-\omega_j^2,\omega_j^2),$$
$$\pi(\omega_j)=\text{Gamma}\Big(3,\frac{\gamma}{\sigma}\Big),\quad j=1,\cdots,q.$$

Hence, the prior distributions of β and Φ are

$$\pi(\beta|\sigma,\lambda)\propto\prod_{i=1}^{p}[\pi(\beta_i|\sigma,\lambda,s_i)\pi(s_i)],$$
$$\pi(\Phi|\sigma,\gamma)\propto\prod_{j=1}^{q}[\pi(\phi_j|\sigma,\gamma,\omega_j)\pi(\omega_j)].$$

The scale parameter σ is assigned a noninformative prior as follows:

$$\pi(\sigma)\propto\frac{1}{\sigma}.$$

The penalty parameters λ and γ are assigned gamma priors as follows:

$$\pi(\lambda)\sim\text{Gamma}(a,b),\quad \pi(\gamma)\sim\text{Gamma}(c,d.)$$

Then, the joint prior distribution of all parameters can be expressed as

$$\pi(\beta,\Phi,\lambda,\gamma,\sigma)=\pi(\beta|\sigma,\lambda)\pi(\Phi|\sigma,\gamma)\pi(\sigma)\pi(\gamma)\pi(\lambda).$$

Incorporating the above prior into the joint hierarchical likelihood (6) results in the joint posterior distribution, as shown as follows:

$$\pi(\beta,\Phi,\sigma,\lambda,\gamma|y,x)\propto\text{L}(\beta,\Phi,\sigma|y,x,\upsilon)\pi(\beta,\Phi,\lambda,\gamma,\sigma).$$

3.2 MCMC algorithm

Let $S=(s_1,\cdots,s_p)$ and $W=(\omega_1,\cdots,\omega_q)$. By combining the prior distributions into the joint hierarchical likelihood (6), we formulate the Bayesian hierarchical model as follows:

$$\begin{aligned}
y^{n\times1}|X,\beta,\Phi,\sigma&\sim\text{L}(\beta,\Phi,\sigma|y,x,\upsilon),\\
\beta|S,\lambda,\sigma&\sim\prod_{i=1}^{p}\text{Uniform}(-s_i^2,s_i^2),\\
\Phi|W,\gamma,\sigma&\sim\prod_{j=1}^{q}\text{Uniform}(-\omega_j^2,\omega_j^2),\\
S|\lambda,\sigma&\sim\prod_{i=1}^{p}\text{Gamma}\Big(3,\frac{\lambda}{\sigma}\Big),\\
W|\gamma,\sigma&\sim\prod_{j=1}^{q}\text{Gamma}\Big(3,\frac{\gamma}{\sigma}\Big),\\
\lambda&\sim\text{Gamma}(a,b),\\
\gamma&\sim\text{Gamma}(c,d),\\
\sigma&\sim\pi(\sigma)\propto\frac{1}{\sigma}.
\end{aligned}\tag{9}$$

In the implementation of the Gibbs sampling algorithm, the involved full conditional posterior distributions of all parameters are presented as follows:

- $\pi(\sigma|y,x,\upsilon,\beta,\Phi,\lambda,\gamma)\sim\text{IGamma}(\delta_1,\delta_2)$,

where $\delta_1=\frac{3(n-q)}{2}+3(p+q)$, $\delta_2=\sum_{t=q+1}^{n}\Big(\frac{e_t^2}{2\theta_2\upsilon_t}+\upsilon_t\Big)+\lambda\sum_{i=1}^{p}s_i+\gamma\sum_{j=1}^{q}\omega_j$, $e_t=\tilde{y}_t-\tilde{x}_t^T\beta-\theta_1\upsilon_t$, $\tilde{y}_t=y_t-\sum_{j=1}^{q}\phi_jy_{t-j}$, and $\tilde{x}_t=x_t-\sum_{j=1}^{q}\phi_jx_{t-j}$, $t=q+1,\cdots,n$.

- $\pi(\lambda|y,x,\upsilon,\beta,\Phi,\sigma,\gamma)\sim\text{Gamma}\big(3p+a,\ b+\sum_{i=1}^{p}s_i/\sigma\big)$.

- $\pi(\gamma|y,x,\upsilon,\beta,\Phi,\sigma,\lambda)\sim\text{Gamma}\big(3q+c,\ d+\sum_{j=1}^{q}\omega_i/\sigma\big)$.

- $\pi(\upsilon_t|y,x,\beta,\Phi,\sigma,\lambda,\gamma)\sim\text{GIG}\Big(\frac{1}{2},\ \frac{\eta_t^2}{\theta_2\sigma},\ \frac{\theta_1^2+2\theta_2}{\theta_2\sigma}\Big)$,

where $\eta_t=\tilde{y}_t-\tilde{x}_t^T\beta,\ t=q+1,\cdots,n$, and $\text{GIG}(\lambda,\chi,\psi)$ denotes the generalized inverse Gaussian distribution with index λ and scale parameters $\chi>0,\psi>0$.

- $\pi(\beta|y,x,\upsilon,\Phi,\lambda,\gamma,\sigma,S,W)\propto\prod_{i=1}^{p}\text{I}(|\beta_i|<s_i^2)\cdot\exp\Big\{-\sum_{t=q+1}^{n}\frac{(y_t-x_t^T\beta-\sum_{j=1}^{q}\phi_j(y_{t-j}-x_{t-j}^T\beta)-\theta_1\upsilon_t)^2}{2\theta_2\sigma\upsilon_t}\Big\}$.

Equivalently,

$$\pi(\beta|y,x,\upsilon,\Phi,\sigma,\lambda,\gamma)\sim N(\beta^*,B^*)\prod_{i=1}^{p}\text{I}(|\beta_i|<s_i^2),$$

where $\beta^*=B^*\Big(\sum_{t=q+1}^{n}\frac{\tilde{x}_t\tilde{\tilde{y}}_t}{\theta_2\sigma\upsilon_t}\Big)$, $B^*=\Big(\sum_{t=q+1}^{n}\frac{\tilde{x}_t\tilde{x}_t^T}{\theta_2\sigma\upsilon_t}\Big)^{-1}$, and $\tilde{\tilde{y}}_t=\tilde{y}_t-\theta_1\upsilon_t,\ t=q+1,\cdots,n$.

- $\pi(\Phi|y,x,\upsilon,\beta,\lambda,\gamma,\sigma,S,W)\propto\prod_{j=1}^{q}\text{I}(|\phi_j|<\omega_j^2)\cdot\exp\Big\{-\sum_{t=q+1}^{n}\frac{(y_t-x_t^T\beta-\sum_{j=1}^{q}\phi_j(y_{t-j}-x_{t-j}^T\beta)-\theta_1\upsilon_t)^2}{2\theta_2\sigma\upsilon_t}\Big\}$.

Equivalently,

$$\pi(\Phi|y,x,\upsilon,\beta,\sigma,\lambda,\gamma)\sim N(\psi,\Psi)\prod_{j=1}^{q}\text{I}(|\varphi_j|<\omega_j^2),$$

where $\psi = \Psi\Big(\sum_{t=q+1}^{n} \frac{R_t \tilde{r}_t}{\theta_2 \sigma \upsilon_t}\Big)$, $\Psi = \Big(\sum_{t=q+1}^{n} \frac{R_t R_t^T}{\theta_2 \sigma \upsilon_t}\Big)^{-1}$, $R_t^T = (r_{t-1}, \cdots, r_{t-q})$, $r_{t-j} = y_{t-j} - x_{t-j}^T \beta$, and $\tilde{r}_t = r_t - \theta_1 \upsilon_t, j = 0, \cdots, q, t = q+1, \cdots, n$.

- $\pi(S|y, X, \upsilon, \beta, \sigma, \Phi, \lambda, \gamma) \sim \prod_{i=1}^{p} \text{Exp}(\lambda/\sigma) I\{s_i > |\beta_i|^{1/2}\}$.
- $\pi(W|y, x, \upsilon, \beta, \sigma, \Phi, \lambda, \gamma) \sim \prod_{j=1}^{q} \text{Exp}(\gamma/\sigma) I\{\omega_j > |\phi_j|^{1/2}\}$.

Remark 1. β and Φ are sampled from truncated multivariate normal distributions.

Remark 2. s_i is generated from the left-truncated exponential distribution through the inversion method, which can be implemented in two steps as follows:

(i) generate $s_i^* \sim \text{Exp}(\lambda/\sigma)$,
(ii) generate $s_i = s_i^* + |\beta_i|^{1/2},\ i = 1, \cdots, p$.

Remark 3. ω_j is generated from the left-truncated exponential distribution in the same manner as in Remark 2.

By sampling repeatedly from the above full conditional distributions, we obtain a series of MCMC samples of the parameters to conduct posterior inference.

4. BAYESIAN ADAPTIVE $L_{1/2}$ PENALIZED QR (BA$L_{1/2}$)

4.1 Prior distributions

Let $\Lambda = (\lambda_1, \cdots, \lambda_p)$ and $\Gamma = (\gamma_1, \cdots, \gamma_q)$. By forcing a distinct penalty parameter on each regression coefficient, BA$L_{1/2}$ can result in more efficient estimation than B$L_{1/2}$. We impose adaptive $L_{1/2}$ penalization priors on the regression coefficients and AR parameters as follows:

$$\begin{aligned}
\pi(\beta|\sigma, \Lambda) &= \prod_{i=1}^{p} \pi(\beta_i|\sigma, \lambda_i),\\
\pi(\beta_i|\sigma, \lambda_i) &= \frac{\lambda_i^2}{4\sigma^2} \exp\Big\{-\frac{\lambda_i}{\sigma}|\beta_i|^{1/2}\Big\};\\
\pi(\Phi|\sigma, \Gamma, \zeta) &= \prod_{j=1}^{q} \pi(\phi_j|\sigma, \gamma),\\
\pi(\phi_j|\sigma, \gamma_j) &= \frac{\gamma_j^2}{4\sigma^2} \exp\Big\{-\frac{\gamma_j}{\sigma}|\phi_j|^{1/2}\Big\}.
\end{aligned}$$

Similarly, the priors of β_i and ϕ_j can be decomposed as follows:

$$\begin{aligned}
\pi(\beta_i|\sigma, \lambda_i) &= \int_0^\infty \pi(\beta_i|\lambda_i, \sigma, s_i)\pi(s_i) ds_i,\\
\pi(\phi_j|\sigma, \gamma_j) &= \int_0^\infty \pi(\phi_j|\gamma_j, \sigma, \omega_j)\pi(\omega_j) d\omega_j,
\end{aligned}$$

where

$$\begin{aligned}
\pi(\beta_i|\lambda_i, \sigma, s_i) &= \text{Uniform}(-s_i^2, s_i^2),\\
\pi(s_i|\xi) &= \text{Ga}\Big(3, \frac{\lambda_i}{\sigma}\Big),\ i = 1, \cdots, p;\\
\pi(\phi_j|\gamma_j, \sigma, \omega_j) &= \text{Uniform}(-\omega_j^2, \omega_j^2),\\
\pi(\omega_j|\zeta) &= \text{Ga}\Big(3, \frac{\gamma_j}{\sigma}\Big),\ j = 1, \cdots, q.
\end{aligned}$$

Hence, the priors of β, Φ, and σ are

$$\begin{aligned}
\pi(\beta|\sigma, \lambda) &\propto \prod_{i=1}^{p} [\pi(\beta_i|\sigma, \lambda_i, s_i)\pi(s_i)],\\
\pi(\Phi|\sigma, \gamma) &\propto \prod_{j=1}^{q} [\pi(\phi_j|\sigma, \gamma_j, \omega_j)\pi(\omega_j)],\\
\pi(\sigma) &\propto \frac{1}{\sigma}.
\end{aligned}$$

The priors of Λ and Γ are

$$\begin{aligned}
\pi(\Lambda) &= \prod_{i=1}^{p} \pi(\lambda_i),\ \pi(\lambda_i) \sim \text{Gamma}(a_i, b_i),\\
\pi(\Gamma) &= \prod_{j=1}^{q} \pi(\gamma_j),\ \pi(\gamma_j) \sim \text{Gamma}(c_j, d_j).
\end{aligned}$$

Likewise, the joint posterior distribution can be written as

$$\begin{aligned}
&\pi(\beta, \Phi, \sigma, \lambda, \gamma|y, x)\\
&\propto \text{L}(\beta, \Phi, \sigma|y, x, \upsilon)\pi(\beta|\sigma, \Lambda)\pi(\Phi|\sigma, \Gamma)\pi(\sigma)\pi(\Lambda)\pi(\Gamma).
\end{aligned} \tag{10}$$

4.2 MCMC algorithm

On the basis of the prior specification, we have the hierarchical representation that is similar to (9) except that λ and γ are replaced by λ_i and γ_j, respectively. Consequently, the full conditional distributions of the parameters are similar to those of Section 3.2 except the followings:

- $\pi(\sigma|y, x, \upsilon, \beta, \Phi, \lambda, \gamma) \sim \text{IGamma}(\delta_1, \delta_2)$,

where $\delta_1 = \frac{3(n-q)}{2} + 3(p+q)$, $\delta_2 = \sum_{t=q+1}^{n}\Big(\frac{e_t^2}{2\theta_2 \upsilon_t} + \upsilon_t\Big) + \sum_{i=1}^{p} \lambda_i s_i + \sum_{j=1}^{q} \gamma_j \omega_j$, and other terms are the same as before.

- $\pi(\lambda_i|y, x, \upsilon, \beta, \Phi, \sigma, \Gamma) \sim \text{Gamma}\big(3 + a_i,\ b_i + s_i/\sigma\big)$, $i = 1, \cdots, p$.
- $\pi(\gamma_j|y, x, \upsilon, \beta, \Phi, \sigma, \Lambda) \sim \text{Gamma}\big(3 + c_j,\ d_j + \omega_j/\sigma\big)$, $j = 1, \cdots, q$.
- $\pi(S|y, X, \upsilon, \beta, \sigma, \Phi, \Lambda, \Gamma) \sim \prod_{i=1}^{p} \text{Exp}(\lambda_i/\sigma) I\{s_i > |\beta_i|^{1/2}\}$.
- $\pi(W|y, x, \upsilon, \beta, \sigma, \Phi, \Lambda, \Gamma) \sim \prod_{j=1}^{q} \text{Exp}(\gamma_j/\sigma) I\{\omega_j > |\phi_j|^{1/2}\}$.

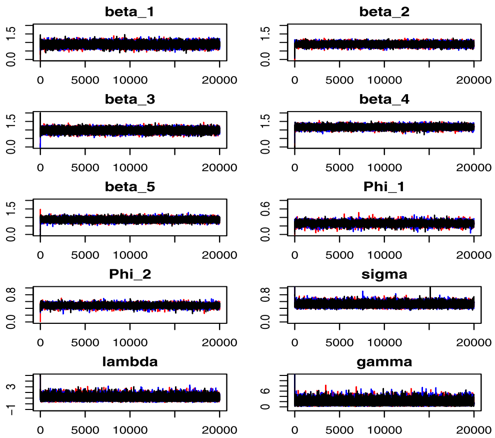

Figure 1. MCMC chains starting from different initial values of $BL_{1/2}$ QR for Model 1. Notes: The red line denotes the first setting; The blue line denotes the second setting; The black line denotes the third setting.

5. SIMULATION STUDY

In this section, simulation studies are conducted to demonstrate the empirical performance of the proposed methods. We generate 50 datasets from model (1) with a sample size of $n = 100$ for three cases as follows:

Model 1: dense case, $\beta = (1, 1, 1, 1, 1)^T$, $\Phi = (0.4, 0.4)^T$.

Model 2: sparse case, $\beta = (1, 1, 0, 0, 0)^T$, $\Phi = (0.5, 0, -0.7, 0)^T$.

Model 3: extremely sparse case, $\beta = (\beta_{1:2}^T, \beta_{3:20}^T)^T$, $\Phi = (0.5, 0, -0.7, 0)^T$, where $\beta_{1:2} = (1, 1)^T, \beta_{3:20} = (0, \cdots, 0)^T$, and $\beta_{a:b}$ denotes the set $\{\beta_j; j = a, \cdots, b\}$.

In the three models, the elements of x_t are independently generated from the standard normal distribution. Model error η_t is generated from the following distributions: (i) $N(0, 1)$, (ii) $t(3)$ (t distribution with three degree of freedom), and (iii) $\chi^2(2)$ (Chi-square distribution with two degree of freedom). Three quantiles $\tau = (0.25, 0.50, 0.75)$ are considered for each case. The hyperparameters of the prior distributions are set to be $a = b = c = d = 0.1$ for $\mathrm{B}L_{1/2}$ and $a_i = b_i = c_j = d_j = 0.1$ for $\mathrm{BA}L_{1/2}$. To guide the convergence of the MCMC algorithm, we consider three different initial values: Initial 1: $\{\beta = (-3, -3, -3, -3, -3)^T$, $\Phi = (0.6, 0.2)^T$, $\sigma = 1\}$, Initial 2: $\{\beta = (1, 1, 1, 1, 1)^T$, $\Phi = (0.3, 0.5)^T$, $\sigma = 5\}$, and Initial 3: $\{\beta = (3, 3, 3, 3, 3)^T$, $\Phi = (0.5, 0.3)^T$, $\sigma = 10\}$, and run three parallel chains with these initial values in the setting of Model 1, $\eta_t \sim t(3)$, $\tau = 0.5$, and $\mathrm{B}L_{1/2}$ penalty. Figure 1 shows that the three MCMC chains starting from the different initial values mix rapidly within hundreds of iterations, thereby indicating a quick convergence of the MCMC algorithm regardless of initial values. To be conservative, in each of the following settings, we run the Gibbs sampling algorithm 20,000 times,

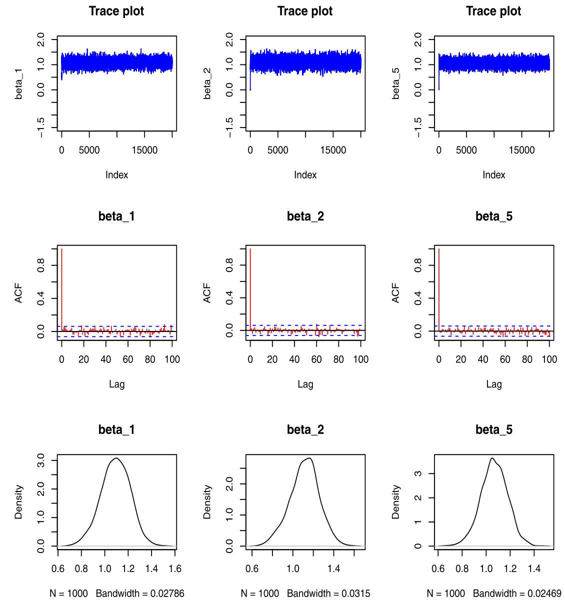

Figure 2. The first row displays one trace plot of 20,000 Gibbs samples of $BAL_{1/2}$ penalized estimation results, the second and third rows represent the ACF plots and density plots of 1,000 sampled values at the median in Model 1 of the simulation.

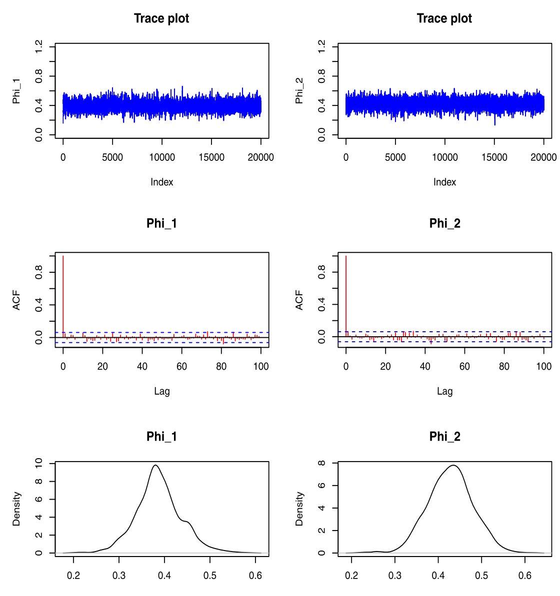

Figure 3. The first row displays one trace plot of 20,000 Gibbs samples of $BAL_{1/2}$ penalized estimation results, the second and third rows represent the ACF plots and density plots of 1,000 sampled values at the median in Model 1 of the simulation.

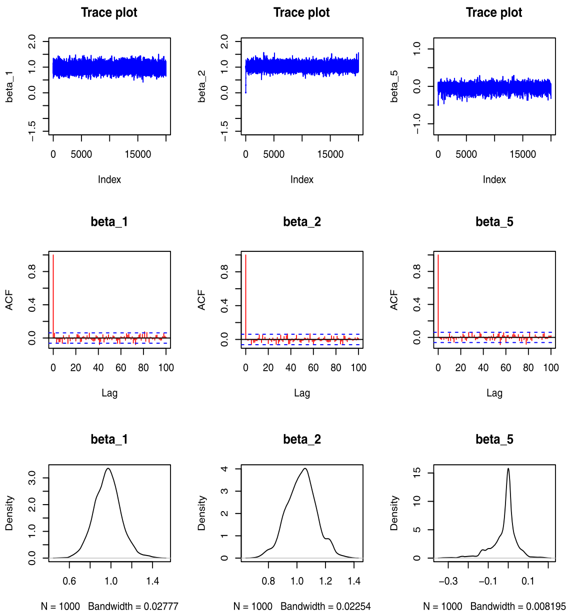

Figure 4. The first row displays one trace plot of 20,000 Gibbs samples of $BAL_{1/2}$ penalized estimation results, the second and third rows represent the ACF plots and density plots of 1,000 sampled values at the median in Model 2 of the simulation.

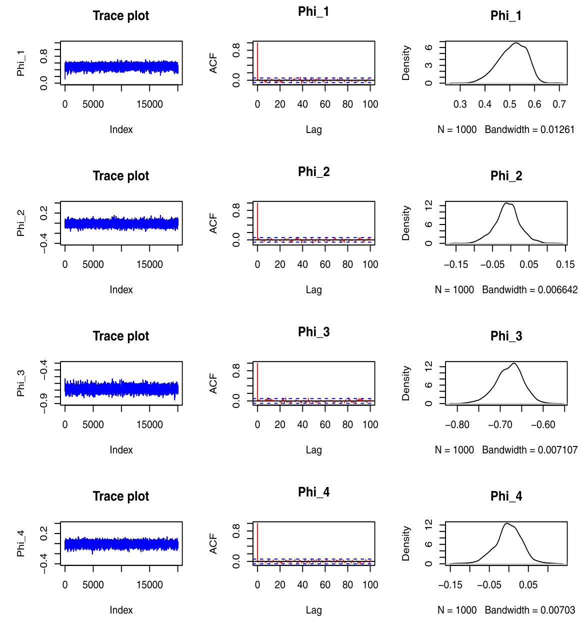

Figure 5. The first column displays one trace plot of 20,000 Gibbs samples of $BAL_{1/2}$ penalized estimation results, the second and third columns represent the ACF plots and density plots of 1,000 sampled values at the median in Model 2 of the simulation.

delete 10,000 burn-ins, and preserve the remaining 10,000 samples to generate 1,000 posterior samples with a step size of 10 for posterior inference. The results are summarized based on 50 replications.

For **Model 1**, the averaged estimation values (Est.), standard error estimates (St.d), and the lower (C.L.) and upper (U.L.) bounds of the 95% credible intervals of β and Φ are reported in Tables 1 and 2. The trace plots of 20,000 Gibbs samples, ACF plots, and the density plots of 1,000 posterior samples in one of the MCMC iterations of regression coefficients β_1, β_2, and β_5 and AR parameters ϕ_1 and ϕ_2 in the setting of $\eta_t \sim t(3)$, $\tau = 0.5$, and $\mathrm{BA}L_{1/2}$ are presented in Figures 2 and 3. The MCMC chains rapidly converge to their stationary distributions, and their posterior distributions are approximately normal. Tables 1 and 2 indicate that $\mathrm{B}L_{1/2}$ and $\mathrm{BA}L_{1/2}$ can estimate parameters accurately for different error distributions over the three quantiles. At $\tau = 0.5$, $\mathrm{B}L_{1/2}$ and $\mathrm{BA}L_{1/2}$ provide superior estimation results with smaller bias and Std for most of the parameters under symmetric errors ($N(0,1)$ and t_3) than those under asymmetric error ($\chi^2(2)$). The estimation results under normal errors in general perform the best at the three quantiles.

For **Model 2**, the main goal is to examine whether $\mathrm{B}L_{1/2}$ and $\mathrm{BA}L_{1/2}$ can specify zero coefficients exactly for a sparse case. The estimation results are reported in Tables 3 and 4. The trace plots of 20,000 Gibbs samples, ACF plots, and the density plots of 1,000 posterior samples in one of the MCMC chains of regression coefficients β_1, β_2, and β_5 and AR parameter Φ in the setting of $\eta_t \sim t_3$, $\tau = 0.5$, and $\mathrm{BA}L_{1/2}$ are presented in Figures 4 and 5. The MCMC chains of regression coefficients rapidly converge to their stationary distributions. In particular, for zero coefficients β_3, β_4, and β_5, their MCMC samples converge to zero stationarily. Tables 3 and 4 indicate that $\mathrm{B}L_{1/2}$ and $\mathrm{BA}L_{1/2}$ can specify significant covariates and zero-valued coefficients accurately regardless of the error distributions and quantile levels. A comparison of the estimation results of $\mathrm{B}L_{1/2}$ and $\mathrm{BA}L_{1/2}$ indicates that the latter estimates zero coefficients more exactly with smaller bias and St.d and shorter credible interval of covering zero than the former does. The proportions of the zero coefficients in β that were correctly identified by $\mathrm{B}L_{1/2}$ at $\tau = (0.25, 0.5, 0.75)$ are (0.991, 0.990, 0.990), (0.991, 0.987, 0.985), and (0.994, 0.985, 0.967) when $\eta_t \sim N(0,1), t(3)$, and $\chi^2(2)$, respectively. In comparison, those identified by $\mathrm{BA}L_{1/2}$ at $\tau = (0.25, 0.5, 0.75)$ are (0.993, 0.996, 0.998), (0.994, 0.996, 0.994), and (0.998, 0.991, 0.983) when $\eta_t \sim N(0,1), t(3)$, and $\chi^2(2)$, respectively.

For **Model 3**, we examine whether the proposed methods can identify zero coefficients exactly for an extremely sparse case. The estimation results of regression coefficients $\beta_1, \beta_2, \beta_5, \beta_{11}$, and β_{18} and AR parameters ϕ_1, ϕ_2, ϕ_3, and ϕ_4 are reported in Tables 5 and 6. The trace plots, ACF plots, and the density plots of the posterior samples of the parameters are similar to those of Model 2 and not reported.

Table 1. Estimation results of $BL_{1/2}$ in **Model 1**

Error	τ	Est.	$\beta_1=1$	$\beta_2=1$	$\beta_3=1$	$\beta_4=1$	$\beta_5=1$	$\phi_1=0.4$	$\phi_2=0.4$
$N(0,1)$	0.25	Est.	0.983	0.981	1.007	1.000	1.004	0.384	0.395
		St.d	0.084	0.108	0.113	0.118	0.106	0.112	0.112
		95% C.L.	0.807	0.740	0.818	0.777	0.786	0.222	0.185
		95% C.U.	1.107	1.201	1.222	1.215	1.216	0.625	0.582
	0.50	Est.	0.992	0.991	1.012	0.957	0.974	0.399	0.350
		St.d	0.106	0.123	0.106	0.118	0.115	0.120	0.105
		95% C.L.	0.795	0.777	0.808	0.731	0.713	0.151	0.115
		95% C.U.	1.159	1.213	1.174	1.114	1.117	0.574	0.534
	0.75	Est.	0.963	0.992	0.980	1.004	0.968	0.421	0.357
		St.d	0.116	0.119	0.129	0.126	0.117	0.106	0.115
		95% C.L.	0.766	0.761	0.757	0.742	0.681	0.248	0.212
		95% C.U.	1.192	1.217	1.199	1.243	1.147	0.622	0.565
t_3	0.25	Est.	0.991	0.989	0.989	0.989	0.997	0.411	0.359
		St.d	0.145	0.115	0.152	0.138	0.139	0.076	0.077
		95% C.L.	0.707	0.803	0.750	0.718	0.791	0.270	0.196
		95% C.U.	0.186	1.190	1.335	1.258	1.252	0.529	0.489
	0.50	Est.	0.996	0.987	0.983	0.976	0.993	0.389	0.378
		St.d	0.116	0.136	0.119	0.097	0.114	0.071	0.056
		95% C.L.	0.834	0.741	0.801	0.770	0.800	0.251	0.278
		95% C.U.	1.289	1.254	1.257	1.157	1.178	0.497	0.461
	0.75	Est.	1.025	0.988	0.975	0.962	0.940	0.378	0.375
		St.d	0.169	0.135	0.113	0.139	0.145	0.096	0.099
		95% C.L.	0.748	0.740	0.770	0.691	0.602	0.236	0.180
		95% C.U.	1.351	1.218	1.188	1.180	1.112	0.586	0.540
χ^2_2	0.25	Est.	0.972	1.015	0.973	0.967	0.971	0.406	0.397
		St.d	0.098	0.075	0.090	0.094	0.105	0.045	0.043
		95% C.L.	0.783	0.879	0.830	0.775	0.785	0.345	0.315
		95% C.U.	1.151	1.159	1.147	1.126	1.151	0.499	0.471
	0.50	Est.	0.962	0.939	0.951	0.990	0.960	0.437	0.375
		St.d	0.161	0.151	0.168	0.165	0.174	0.094	0.087
		95% C.L.	0.646	0.656	0.660	0.681	0.679	0.279	0.209
		95% C.U.	1.217	1.215	1.280	1.318	1.184	0.656	0.546
	0.75	Est.	0.958	0.946	0.999	0.906	0.975	0.401	0.338
		St.d	0.293	0.299	0.261	0.296	0.255	0.136	0.137
		95% C.L.	0.441	0.391	0.450	0.397	0.450	0.140	0.137
		95% C.U.	1.379	1.442	1.418	1.411	1.458	0.640	0.585

Tables 5 and 6 indicate that B$L_{1/2}$ and BA$L_{1/2}$ can specify significant covariates and zero-valued coefficients accurately for the extremely sparse case. The proportions of the zero coefficients in β that were correctly identified by B$L_{1/2}$ at $\tau = (0.25, 0.5, 0.75)$ are (0.960, 0.979, 0.968), (0.965, 0.972, 0.954), and (0.984, 0.966, 0.900) when $\eta_t \sim N(0,1), t(3)$, and $\chi^2(2)$, respectively. In comparison, those identified by BA$L_{1/2}$ at $\tau = (0.25, 0.5, 0.75)$ are (0.977, 0.983, 0.978), (0.947, 0.974, 0.961), and (0.978, 0.954, 0.901) when $\eta_t \sim N(0,1), t(3)$, and $\chi^2(2)$, respectively.

To compare the estimation results of the proposed methods with Bayesian LASSO (BLASSO) and Bayesian adaptive LASSO (BALASSO) penalties, we provide an evaluation quantity MMSE, which is defined as follows:

$$\text{MMSE} = \text{Mean}\Big((\hat{\theta} - \theta^{true})^T (\hat{\theta} - \theta^{true})\Big), \tag{11}$$

where $\theta = (\beta^T, \Phi^T)^T$ and the Mean$(\cdot)$ denotes the average value of the mean square errors based on 50 replications. The MMSE values of Models 1–3 under B$L_{1/2}$, BA$L_{1/2}$, BLASSO, and BALASSO are presented in Table 7.

BLASSO-based methods perform slightly better than B$L_{1/2}$-based methods in the dense case (Model 1). However, in the sparse and extremely sparse cases (Models 2 and 3), the proposed B$L_{1/2}$-based methods perform much better than BLASSO-type methods with smaller MMSE values for most of cases over the three quantiles. Regarding computational efficiency, the proposed B$L_{1/2}$-based methods are generally faster than BLASSO-based methods. Taking the setting of Model 2: $\eta_t \sim t(3)$ and $\tau = 0.5$ as an example, the computing times of B$L_{1/2}$ and BA$L_{1/2}$ for completing one replication are 4.589 and 4.697 minutes, respectively, whereas those of BLASSO and BALASSO are 5.132 and 5.241 minutes, respectively.

Table 2. Estimation results of $BAL_{1/2}$ in **Model 1**

Error	τ	Est.	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_3 = 1$	$\beta_4 = 1$	$\beta_5 = 1$	$\phi_1 = 0.4$	$\phi_2 = 0.4$
$N(0,1)$	0.25	Est.	0.996	1.002	0.995	1.016	0.973	0.400	0.382
		St.d	0.109	0.106	0.106	0.134	0.095	0.108	0.107
		95% C.L.	0.778	0.831	0.800	0.779	0.803	0.233	0.134
		95% C.U.	1.189	1.236	1.170	1.249	1.127	0.675	0.548
	0.50	Est.	0.994	0.999	1.002	0.995	0.990	0.409	0.346
		St.d	0.121	0.109	0.117	0.115	0.091	0.111	0.100
		95% C.L.	0.783	0.817	0.803	0.720	0.811	0.224	0.173
		95% C.U.	1.209	1.184	1.253	1.227	1.135	0.598	0.526
	0.75	Est.	0.981	1.019	1.001	1.020	1.003	0.378	0.397
		St.d	0.116	0.103	0.122	0.135	0.091	0.114	0.110
		95% C.L.	0.799	0.811	0.765	0.790	0.826	0.209	0.201
		95% C.U.	1.219	1.196	1.250	1.288	1.169	0.589	0.558
t_3	0.25	Est.	0.965	0.983	0.971	1.002	0.969	0.390	0.378
		St.d	0.166	0.148	0.173	0.161	0.149	0.085	0.088
		95% C.L.	0.608	0.706	0.631	0.692	0.673	0.278	0.218
		95% C.U.	1.250	1.307	1.241	1.368	1.306	0.563	0.520
	0.50	Est.	0.983	0.996	0.965	1.001	0.975	0.367	0.400
		St.d	0.137	0.124	0.119	0.124	0.118	0.067	0.075
		95% C.L.	0.696	0.760	0.700	0.759	0.725	0.219	0.266
		95% C.U.	1.199	1.202	1.158	1.214	1.134	0.461	0.533
	0.75	Est.	0.979	1.024	0.982	0.966	0.972	0.390	0.373
		St.d	0.153	0.158	0.159	0.154	0.158	0.089	0.087
		95% C.L.	0.665	0.683	0.627	0.669	0.664	0.217	0.223
		95% C.U.	1.244	1.279	1.248	1.211	1.238	0.548	0.502
χ^2_2	0.25	Est.	0.967	1.028	1.003	1.005	0.980	0.410	0.394
		St.d	0.087	0.095	0.082	0.091	0.098	0.050	0.050
		95% C.L.	0.778	0.838	0.862	0.860	0.800	0.316	0.299
		95% C.U.	1.092	1.171	1.144	1.188	1.165	0.505	0.501
	0.50	Est.	0.953	0.951	0.958	0.970	0.970	0.407	0.404
		St.d	0.167	0.150	0.168	0.177	0.153	0.079	0.075
		95% C.L.	0.574	0.725	0.653	0.621	0.664	0.271	0.268
		95% C.U.	1.276	1.227	1.278	1.310	1.260	0.565	0.567
	0.75	Est.	0.953	0.971	0.917	1.003	1.016	0.394	0.345
		St.d	0.302	0.309	0.311	0.269	0.251	0.113	0.123
		95% C.L.	0.345	0.368	0.388	0.487	0.489	0.213	0.183
		95% C.U.	1.411	1.453	1.588	1.400	1.450	0.635	0.516

The computations are implemented using statistical software R3.5.2 in a laptop [Dell Intel(R) Core(TM) i7-6600U CPU]. The R codes are freely available upon request.

6. REAL DATA APPLICATION

In this section, we applied the proposed methods to a real-life study concerning the electricity consumption by residential customers served by the San Diego Gas and Electric Company. The dataset was obtained from Ramanathan (2002) and analyzed by previous studies (e.g., Lim and Oh, 2014; Yoon et al., 2017; Tian et al., 2019). Eighty-seven quarterly observations were recorded in the dataset. The response variable is the electricity consumption, which is measured by the logarithm of the kilowatt-hour (kwh) sales per residential customer (LKWH). The predictor variables are the logarithm of per capita income (LIncome), the logarithm of average price of residential electricity in dollars per kwh (LPrice), cooling degree days (CDD), and heating degree days (HDD). In this application, we aimed to examine the potential influential factors of electricity consumption at its various quantiles. In particular, we are interested in investigating how customers' income and electricity price affect the consumption of electricity for customers whose consumption level is low or high. Moreover, previous studies (e.g., Yoon et al., 2017) showed that an ordinary linear regression model with i.i.d. errors is inadequate because the model residuals exhibited considerable autocorrelations. The proposed QR model with AR errors perfectly accommodates the abovementioned needs. The $BL_{1/2}$ and $BAL_{1/2}$ procedures were performed for the subsequent inference.

Table 3. Estimation results of $BL_{1/2}$ in **Model 2**

Error	τ	Est.	$\beta_1=1$	$\beta_2=1$	$\beta_3=0$	$\beta_4=0$	$\beta_5=0$	$\phi_1=0.5$	$\phi_2=0$	$\phi_3=-0.7$	$\phi_4=0$
$N(0,1)$	0.25	Est.	1.018	0.987	−0.013	−0.013	0.014	0.504	−0.006	−0.663	0.012
		St.d	0.109	0.087	0.072	0.065	0.055	0.097	0.099	0.121	0.090
		95% C.L.	0.846	0.837	−0.157	−0.168	−0.080	0.348	−0.174	−0.853	−0.144
		95% C.U.	1.291	1.127	0.102	0.112	0.116	0.686	0.158	−0.379	0.235
	0.50	Est.	0.998	0.976	−0.009	−0.022	−0.005	0.477	−0.005	−0.666	−0.021
		St.d	0.085	0.098	0.072	0.079	0.071	0.120	0.100	0.112	0.109
		95% C.L.	0.854	0.784	−0.177	−0.223	−0.140	0.223	−0.198	−0.858	−0.221
		95% C.U.	1.200	1.156	0.089	0.092	0.135	0.692	0.189	−0.486	0.163
	0.75	Est.	0.992	0.998	−0.002	0.005	0.009	0.485	−0.011	−0.674	−0.017
		St.d	0.106	0.100	0.077	0.050	0.075	0.102	0.104	0.123	0.090
		95% C.L.	0.815	0.850	−0.137	−0.100	−0.117	0.330	−0.272	−0.873	−0.213
		95% C.U.	1.210	1.204	0.155	0.086	0.153	0.689	0.135	−0.415	0.145
t_3	0.25	Est.	0.997	0.997	0.002	−0.000	−0.040	0.483	−0.014	−0.666	−0.021
		St.d	0.133	0.147	0.071	0.072	0.071	0.102	0.097	0.099	0.099
		95% C.L.	0.736	0.693	−0.118	−0.144	−0.205	0.279	−0.173	−0.835	−0.217
		95% C.U.	1.192	1.316	0.114	0.141	0.117	0.648	0.199	−0.434	0.139
	0.50	Est.	0.982	0.974	0.016	−0.016	0.014	0.490	−0.014	−0.666	−0.015
		St.d	0.126	0.097	0.083	0.080	0.079	0.088	0.058	0.059	0.077
		95% C.L.	0.716	0.817	−0.137	−0.187	−0.162	0.325	−0.180	−0.760	−0.133
		95% C.U.	1.201	1.175	0.174	0.133	0.185	0.631	0.072	−0.546	0.119
	0.75	Est.	0.982	0.975	0.035	−0.006	0.012	0.468	−0.001	−0.676	−0.018
		St.d	0.122	0.121	0.107	0.087	0.097	0.094	0.080	0.081	0.078
		95% C.L.	0.773	0.773	−0.110	−0.202	−0.156	0.298	−0.167	−0.813	−0.188
		95% C.U.	1.273	1.202	0.231	0.163	0.185	0.650	0.122	−0.510	0.091
χ^2_2	0.25	Est.	0.972	0.975	0.012	0.018	−0.008	0.522	−0.010	−0.690	0.016
		St.d	0.088	0.093	0.059	0.046	0.064	0.038	0.047	0.053	0.033
		95% C.L.	0.779	0.791	−0.080	−0.043	−0.111	0.445	−0.098	−0.784	−0.049
		95% C.U.	1.100	1.174	0.108	0.129	0.151	0.595	0.055	−0.597	0.098
	0.50	Est.	0.962	0.991	−0.018	−0.001	0.001	0.512	−0.011	−0.680	0.010
		St.d	0.137	0.145	0.096	0.088	0.109	0.084	0.054	0.066	0.082
		95% C.L.	0.668	0.705	−0.210	−0.138	−0.274	0.375	−0.125	−0.785	−0.139
		95% C.U.	1.203	1.226	0.181	0.197	0.178	0.668	0.080	−0.570	0.161
	0.75	Est.	0.965	0.906	−0.014	−0.012	0.004	0.473	0.002	−0.669	−0.018
		St.d	0.250	0.185	0.148	0.191	0.174	0.132	0.126	0.128	0.120
		95% C.L.	0.503	0.580	−0.282	−0.385	−0.268	0.290	−0.321	−0.863	−0.245
		95% C.U.	1.434	1.214	0.278	0.304	0.375	0.777	0.215	−0.377	0.257

We centralized the response variable, standardized the covariates, and considered a model as follows:

$$\mathrm{LKWH}_t = \beta_1 \mathrm{LIncome}_t + \beta_2 \mathrm{LPrice}_t + \beta_3 \mathrm{CDD}_t + \beta_4 \mathrm{HDD}_t + \varepsilon_t. \tag{12}$$

We first used the ACF and PACF plots to examine the autocorrelations of the residuals of model (12). Based on the ACF and PACF plots presented in Figure 6, we set $q = 5$. The QQ plots (not reported) of model (12) with AR(5) errors exhibit linearity over various quantile levels, thereby indicating that the proposed model is plausible for fitting the data. We chose τ to be the three quartiles ($\tau = 0.25, 0.5, 0.75$) because they are the most representative statistics to summarize the lower, central, and upper tendencies of the response variable. The QR analysis at these three levels enables us to understand how potential predictors, such as LIncome, LPrice, CDD, and HDD, influence electricity consumption for low-, median-, and high-electricity consumption populations. In certain circumstances when the extreme tails of the response distribution is of interest, quantile levels of $\tau = 0.1$ and 0.9 may also be considered.

The initial values and prior inputs of the parameters were set in the same manner as those of Simulation 1. After checking convergence, we collected 20,000 posterior samples, deleted 10,000 burn-ins, and preserved the remaining 10,000 samples to generate 1,000 posterior samples with a step size of 10. The estimates, standard error estimates, and the lower and upper bounds of the 95% credible intervals of the parameters are presented in Table 8. The trace plots, ACF plots, and the density functions of the regression coefficients at $\tau = 0.5$ and $\mathrm{BA}L_{1/2}$ are presented in Figure 7, and those

Table 4. Estimation results of $BAL_{1/2}$ in **Model 2**

Error	τ	Est.	$\beta_1=1$	$\beta_2=1$	$\beta_3=0$	$\beta_4=0$	$\beta_5=0$	$\phi_1=0.5$	$\phi_2=0$	$\phi_3=-0.7$	$\phi_4=0$
$N(0,1)$	0.25	Est.	1.004	0.989	0.007	−0.000	0.003	0.481	−0.016	−0.661	−0.018
		St.d	0.104	0.087	0.061	0.043	0.057	0.086	0.078	0.125	0.089
		95% C.L.	0.865	0.820	−0.112	−0.100	−0.107	0.304	−0.184	−0.838	−0.207
		95% C.U.	1.170	1.130	0.157	0.106	0.156	0.617	0.106	−0.392	0.119
	0.50	Est.	0.975	0.982	−0.013	0.014	−0.003	0.497	−0.024	−0.650	−0.020
		St.d	0.094	0.097	0.043	0.047	0.044	0.104	0.087	0.105	0.101
		95% C.L.	0.804	0.834	−0.100	−0.038	−0.119	0.284	−0.203	−0.858	−0.177
		95% C.U.	1.147	1.128	0.031	0.135	0.010	0.692	0.129	−0.456	0.181
	0.75	Est.	1.002	0.987	−0.005	0.006	−0.003	0.494	−0.007	−0.675	−0.005
		St.d	0.091	0.101	0.046	0.037	0.043	0.107	0.113	0.095	0.076
		95% C.L.	0.827	0.786	−0.108	−0.057	−0.084	0.258	−0.224	−0.846	−0.148
		95% C.U.	1.154	1.179	0.080	0.083	0.082	0.687	0.208	−0.508	0.147
t_3	0.25	Est.	1.008	0.975	−0.001	−0.001	0.002	0.495	−0.021	−0.664	−0.012
		St.d	0.116	0.135	0.068	0.061	0.067	0.102	0.064	0.086	0.092
		95% C.L.	0.836	0.737	−0.142	−0.087	−0.072	0.302	−0.140	−0.800	−0.179
		95% C.U.	1.165	1.263	0.177	0.160	0.153	0.707	0.074	−0.498	0.173
	0.50	Est.	0.986	0.997	0.009	−0.007	0.009	0.489	−0.009	−0.659	−0.020
		St.d	0.115	0.100	0.065	0.061	0.053	0.088	0.081	0.090	0.087
		95% C.L.	0.727	0.813	−0.077	−0.121	−0.071	0.324	−0.182	−0.818	−0.219
		95% C.U.	1.193	1.186	0.136	0.137	0.100	0.654	0.116	−0.460	0.127
	0.75	Est.	0.985	1.010	−0.008	−0.002	−0.014	0.474	0.007	−0.675	−0.021
		St.d	0.135	0.129	0.067	0.057	0.051	0.088	0.085	0.101	0.071
		95% C.L.	0.714	0.747	−0.153	−0.139	−0.102	0.317	−0.182	−0.865	−0.166
		95% C.U.	1.168	1.270	0.098	0.100	0.070	0.624	0.148	−0.501	0.134
χ^2_2	0.25	Est.	0.999	0.986	0.006	−0.001	0.005	0.512	−0.001	−0.699	0.018
		St.d	0.087	0.096	0.045	0.035	0.039	0.051	0.043	0.052	0.044
		95% C.L.	0.809	0.825	−0.056	−0.060	−0.056	0.434	−0.089	−0.781	−0.042
		95% C.U.	1.170	1.189	0.132	0.078	0.078	0.607	0.074	−0.594	0.135
	0.50	Est.	0.994	0.986	−0.016	0.004	−0.020	0.516	−0.007	−0.681	0.008
		St.d	0.113	0.140	0.072	0.066	0.088	0.083	0.074	0.076	0.072
		95% C.L.	0.789	0.699	−0.166	−0.097	−0.261	0.408	−0.144	−0.825	−0.118
		95% C.U.	1.195	1.214	0.112	0.134	0.081	0.652	0.139	−0.525	0.135
	0.75	Est.	0.967	0.983	0.011	−0.038	−0.026	0.501	−0.011	−0.673	−0.006
		St.d	0.223	0.239	0.110	0.165	0.131	0.138	0.103	0.125	0.132
		95% C.L.	0.600	0.579	−0.193	−0.448	−0.227	0.218	−0.267	−0.890	−0.309
		95% C.U.	1.492	1.469	0.207	0.243	0.275	0.726	0.143	−0.448	0.235

of AR parameters are not reported. The model error was evaluated by the average QR residual sum (AQRRS)

(13)

$$\text{AQRRS} = \frac{1}{n-q}\sum_{t=q+1}^{n}\rho_\tau[(y_t - x_t^T\hat{\beta}) - \sum_{j=1}^{q}\hat{\phi}_j(y_{t-j} - x_{t-j}^T\hat{\beta})].$$

Table 8 shows that the proposed methods identified distinct predictors at each quantile. CDD and HDD show significant positive effects on KLWH over the three quantiles, thereby indicating that a raise in cooling or heating degree days increases the electricity consumption of residential customers regardless of quantile levels. This conclusion is consistent with those of Ramanathan (2002) and Tian et al. (2019). In addition, we obtained the following new observations. The effect of LIncome on KLWH is nonsignificant at $\tau = 0.25$ but positive at $\tau = 0.5$ and 0.75, and this positive effect becomes increasingly pronounced when τ increases from 0.5 to 0.75. This result implies that at a low level of electricity consumption, per capital income hardly affects the amount of electricity consumption. However, at the median to a relatively high level of electricity consumption, high per capital income results in high electricity consumption. On the contrary, LPrice has a significant negative effect on KLWH at $\tau = 0.25$, but the effect becomes increasingly insignificant when τ varies from 0.5 to 0.75. This result indicates that at a low level of electricity consumption, high electricity price induces low electricity consumption. However, at the median to a relatively high level of electricity consumption, the impact of electricity price on its consumption becomes increasingly unimportant. Regarding the AR parameters, ϕ_1, ϕ_4, and ϕ_5 are apparently different from zero over the three quartiles, whereas ϕ_2 and ϕ_3 are not

Table 5. Estimation results of $BL_{1/2}$ in **Model 3**

Error	τ	Est.	$\beta_1=1$	$\beta_2=1$	$\beta_5=0$	$\beta_{11}=0$	$\beta_{18}=0$	$\phi_1=0.5$	$\phi_2=0$	$\phi_3=-0.7$	$\phi_4=0$
$N(0,1)$	0.25	Est.	0.967	0.956	−0.001	−0.002	−0.001	0.434	−0.019	−0.584	−0.055
		St.d	0.116	0.110	0.059	0.073	0.059	0.109	0.098	0.099	0.080
		95% C.L.	0.735	0.790	−0.105	−0.152	−0.133	0.221	−0.240	−0.766	−0.166
		95% C.U.	1.154	1.138	0.113	0.131	0.087	0.617	0.144	−0.402	0.113
	0.50	Est.	0.949	0.973	0.000	−0.003	0.008	0.418	−0.019	−0.575	−0.076
		St.d	0.107	0.081	0.046	0.043	0.050	0.102	0.081	0.091	0.080
		95% C.L.	0.741	0.842	−0.069	−0.094	−0.104	0.190	−0.175	−0.733	−0.276
		95% C.U.	1.144	1.119	0.088	0.070	0.084	0.598	0.151	−0.417	0.049
	0.75	Est.	0.964	0.938	−0.002	0.003	0.001	0.442	−0.017	−0.586	−0.063
		St.d	0.112	0.095	0.064	0.073	0.069	0.113	0.086	0.106	0.090
		95% C.L.	0.739	0.727	−0.099	−0.152	−0.135	0.236	−0.180	−0.789	−0.230
		95% C.U.	1.163	1.091	0.091	0.126	0.107	0.600	0.129	−0.390	0.074
t_3	0.25	Est.	0.869	0.855	0.007	0.005	−0.003	0.431	−0.026	−0.605	−0.082
		St.d	0.189	0.177	0.075	0.064	0.081	0.089	0.077	0.093	0.080
		95% C.L.	0.427	0.583	−0.111	−0.125	−0.188	0.302	−0.184	−0.768	−0.219
		95% C.U.	1.128	1.209	0.185	0.136	0.161	0.627	0.071	−0.412	0.043
	0.50	Est.	0.918	0.892	0.001	0.009	−0.004	0.443	−0.021	−0.622	−0.054
		St.d	0.143	0.125	0.070	0.053	0.051	0.078	0.063	0.072	0.058
		95% C.L.	0.668	0.687	−0.149	−0.077	−0.084	0.295	−0.156	−0.728	−0.169
		95% C.U.	1.156	1.126	0.136	0.139	0.070	0.593	0.100	−0.477	0.033
	0.75	Est.	0.927	0.914	−0.019	−0.006	0.014	0.422	−0.107	−0.612	−0.077
		St.d	0.143	0.143	0.070	0.059	0.074	0.084	0.068	0.078	0.071
		95% C.L.	0.644	0.665	−0.148	−0.102	−0.089	0.308	−0.140	−0.736	−0.205
		95% C.U.	1.142	1.144	0.067	0.140	0.238	0.591	0.112	−0.467	0.060
χ_2^2	0.25	Est.	0.943	0.940	−0.002	−0.002	−0.001	0.503	−0.008	−0.656	−0.002
		St.d	0.125	0.124	0.058	0.049	0.046	0.042	0.047	0.059	0.042
		95% C.L.	0.733	0.646	−0.122	−0.065	−0.065	0.429	−0.081	−0.770	−0.078
		95% C.U.	1.119	1.126	0.093	0.098	0.100	0.575	0.078	−0.567	0.083
	0.50	Est.	0.891	0.833	0.005	0.000	−0.009	0.474	−0.010	−0.638	−0.019
		St.d	0.160	0.164	0.049	0.052	0.084	0.073	0.061	0.080	0.069
		95% C.L.	0.543	0.453	−0.082	−0.088	−0.154	0.320	−0.117	−0.768	−0.146
		95% C.U.	1.113	1.080	0.095	0.103	0.110	0.604	0.113	−0.470	0.110
	0.75	Est.	0.800	0.852	−0.007	−0.027	−0.018	0.432	−0.044	−0.568	−0.090
		St.d	0.271	0.288	0.107	0.152	0.095	0.115	0.099	0.125	0.112
		95% C.L.	0.267	0.335	−0.187	−0.362	−0.261	0.217	−0.244	−0.812	−0.329
		95% C.U.	1.315	1.341	0.190	0.222	0.113	0.655	0.135	−0.291	0.075

different from zero because their 95% credible intervals consistently cover zero. Table 8 also shows that $\mathrm{BA}L_{1/2}$ can shrink nonsignificant coefficients closer to zeros than $\mathrm{B}L_{1/2}$ does.

Several previous studies (Ramanathan, 2002; Lim and Oh, 2014; Yoon et al., 2017; Tian et al., 2019) also analyzed the same dataset. However, they cannot provide the aforementioned insights into the quantile-specific predictor effects. Specifically, Ramanathan (2002) and Yoon et al. (2017) considered a mean regression rather than QR, thereby only revealing the predictor effects on the mean of electricity consumption. Lim and Oh (2014) and Tian et al. (2019) identified either {LIncome, LPrice} or {CDD, HDD} as common nonzero predictors over the quartile levels. However, they did not differentiate quantile-specific influential factors from common nonzero predictors, thereby failing to tell how customers' income and electricity price influence electricity consumption for different customer populations, especially those of low- and high-electricity consumption.

7. CONCLUSION

We considered $\mathrm{B}L_{1/2}$ and $\mathrm{BA}L_{1/2}$ penalized methods for a linear QR model with AR errors. By using the working likelihood of ALD and the GGD priors of the regression parameters, we constructed hierarchical penalized QR models to conduct Bayesian inference. Simulations and a real data example were conducted to demonstrate the proposed methodology. Results show that $\mathrm{B}L_{1/2}$ and $\mathrm{BA}L_{1/2}$ can identify significant regression predictors accurately and outperform BLASSO and BALASSO in terms of estimation/selection accuracy and computational efficiency for sparse models.

Table 6. Estimation results of $BAL_{1/2}$ in **Model 3**

Error	τ	Est.	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_5 = 0$	$\beta_{11} = 0$	$\beta_{18} = 0$	$\phi_1 = 0.5$	$\phi_2 = 0$	$\phi_3 = -0.7$	$\phi_4 = 0$
$N(0,1)$	0.25	Est.	0.969	0.991	−0.000	−0.001	0.002	0.456	−0.020	−0.599	−0.058
		St.d	0.119	0.113	0.050	0.058	0.053	0.094	0.091	0.109	0.095
		95% C.L.	0.769	0.754	−0.068	−0.107	−0.104	0.289	−0.171	−0.773	−0.209
		95% C.U.	1.140	1.201	0.144	0.137	0.070	0.645	0.151	−0.313	0.065
	0.50	Est.	1.006	0.997	0.001	0.004	−0.004	0.458	−0.013	−0.621	−0.037
		St.d	0.111	0.086	0.033	0.042	0.032	0.114	0.069	0.093	0.074
		95% C.L.	0.783	0.854	−0.052	−0.068	−0.048	0.256	−0.156	−0.774	−0.202
		95% C.U.	1.206	1.173	0.061	0.104	0.066	0.688	0.092	−0.425	0.072
	0.75	Est.	0.981	1.005	−0.002	0.007	−0.003	0.473	−0.019	−0.632	−0.033
		St.d	0.112	0.118	0.063	0.054	0.059	0.087	0.085	0.103	0.072
		95% C.L.	0.800	0.789	−0.153	−0.067	−0.134	0.316	−0.198	−0.826	−0.175
		95% C.U.	1.165	1.210	0.058	0.124	0.085	0.647	0.132	−0.417	0.074
t_3	0.25	Est.	1.004	0.984	0.002	0.002	0.021	0.449	−0.020	−0.624	−0.050
		St.d	0.148	0.146	0.068	0.062	0.083	0.099	0.076	0.100	0.070
		95% C.L.	0.687	0.745	−0.125	−0.108	−0.097	0.245	−0.184	−0.804	−0.190
		95% C.U.	1.255	1.268	0.158	0.126	0.193	0.625	0.069	−0.412	0.064
	0.50	Est.	0.991	0.978	0.002	0.013	0.005	0.453	0.000	−0.639	−0.036
		St.d	0.115	0.112	0.072	0.057	0.045	0.077	0.047	0.082	0.076
		95% C.L.	0.813	0.784	−0.143	−0.081	−0.080	0.328	−0.088	−0.778	−0.186
		95% C.U.	1.213	1.162	0.132	0.162	0.096	0.601	0.105	−0.472	0.067
	0.75	Est.	0.973	0.991	0.001	−0.004	−0.008	0.428	−0.006	−0.634	−0.060
		St.d	0.144	0.143	0.047	0.100	0.064	0.094	0.078	0.102	0.086
		95% C.L.	0.664	0.709	−0.085	−0.125	−0.166	0.231	−0.211	−0.826	−0.245
		95% C.U.	1.205	1.209	0.077	0.310	0.095	0.561	0.128	−0.430	0.059
χ^2_2	0.25	Est.	0.978	0.972	0.005	−0.001	0.000	0.503	−0.014	−0.665	−0.005
		St.d	0.110	0.097	0.042	0.062	0.037	0.051	0.036	0.052	0.034
		95% C.L.	0.733	0.799	−0.081	−0.112	−0.075	0.417	−0.080	−0.756	−0.059
		95% C.U.	1.132	1.184	0.084	0.127	0.068	0.602	0.044	−0.563	0.052
	0.50	Est.	0.912	1.002	0.012	−0.007	0.006	0.491	−0.024	−0.643	−0.014
		St.d	0.180	0.148	0.056	0.085	0.085	0.064	0.055	0.075	0.053
		95% C.L.	0.588	0.767	−0.067	−0.177	−0.173	0.388	−0.144	−0.804	−0.115
		95% C.U.	1.251	1.288	0.167	0.202	0.168	0.627	0.044	−0.478	0.082
	0.75	Est.	0.934	0.959	−0.009	0.013	−0.027	0.464	−0.036	−0.573	−0.061
		St.d	0.235	0.259	0.122	0.110	0.187	0.149	0.094	0.145	0.127
		95% C.L.	0.421	0.561	−0.314	−0.188	−0.417	0.215	−0.230	−0.804	−0.345
		95% C.U.	1.417	1.438	0.226	0.262	0.312	0.772	0.122	−0.311	0.132

There are several directions for further research. First, the proposed model can be extended to incorporate quadratic and interaction terms of $x_{t1}, \cdots, x_{tp}$ or AR errors of higher order without difficulty. However, generalizing the current model to its nonlinear version $y_t = f(x_t^T \beta) + \varepsilon_t$ with a known/unknown function $f(\cdot)$ and a more sophisticated error structure, such as ARMA(p, q), is nontrivial. Considering the nonlinearity of $f(\cdot)$ and the complex structure of ε_t, the involved posterior distributions and sampling schemes become complicated. Developing a valid Bayesian regularized procedure (e.g., Feng et al., 2015, 2017; Wang et al., 2019) to analyze this generalized model is an important topic but requires further investigation. Second, the proposed model assumes that the regression coefficients are time invariant. In substantive research, the predictor effects may vary over time. Thus, an extension of introducing varying coefficients into the current model framework is of scientific interest and worthy of future study. Finally, we assumed that the observed data are complete. However, missing data are common in practice. How to manage missing data, especially missing not at random data, in the context of QR models with dependent errors is an interesting but challenging problem. Nevertheless, the aforementioned advances require substantial efforts in the future.

ACKNOWLEDGEMENTS

The work is supported by China Postdoctoral Science Foundation (No. 2017M610156), National Natural Science Foundation of China (No. 71473280), Research Grant Council of the Hong Kong Special Administration Region (Nos. 14303017 and 14301918), and the Direct Grants of the Chinese University of Hong Kong.

Table 7. MMSE values of $BL_{1/2}$- and BLASSO-type methods

Model	Error	τ	$\text{BL}_{1/2}$	$\text{BAL}_{1/2}$	BLASSO	BALASSO
Model 1	$N(0,1)$	0.25	0.081 (0.070)	0.084 (0.057)	0.080 (0.038)	0.102 (0.057)
		0.50	0.094 (0.065)	0.086 (0.042)	0.072 (0.040)	0.081 (0.041)
		0.75	0.102 (0.066)	0.091 (0.049)	0.088 (0.043)	0.088 (0.062)
	t_3	0.25	0.108 (0.071)	0.144 (0.072)	0.095 (0.050)	0.137 (0.080)
		0.50	0.077 (0.045)	0.090 (0.052)	0.084 (0.044)	0.091 (0.056)
		0.75	0.125 (0.103)	0.140 (0.114)	0.083 (0.051)	0.119 (0.086)
	$\chi^2(2)$	0.25	0.050 (0.033)	0.048 (0.031)	0.115 (0.069)	0.054 (0.040)
		0.50	0.160 (0.102)	0.151 (0.090)	0.131 (0.092)	0.132 (0.067)
		0.75	0.443 (0.251)	0.453 (0.230)	0.225 (0.122)	0.558 (0.309)
Model 2	$N(0,1)$	0.25	0.075 (0.061)	0.066 (0.054)	0.193 (0.111)	0.076 (0.043)
		0.50	0.084 (0.061)	0.068 (0.049)	0.100 (0.066)	0.085 (0.054)
		0.75	0.080 (0.085)	0.063 (0.042)	0.219 (0.112)	0.085 (0.058)
	t_3	0.25	0.096 (0.079)	0.076 (0.061)	0.128 (0.078)	0.089 (0.059)
		0.50	0.068 (0.037)	0.066 (0.056)	0.066 (0.042)	0.062 (0.045)
		0.75	0.089 (0.057)	0.077 (0.064)	0.143 (0.067)	0.090 (0.053)
	$\chi^2(2)$	0.25	0.036 (0.027)	0.031 (0.022)	0.164 (0.102)	0.029 (0.019)
		0.50	0.091 (0.065)	0.074 (0.049)	0.099 (0.069)	0.107 (0.071)
		0.75	0.258 (0.147)	0.226 (0.142)	0.351 (0.170)	0.301 (0.204)
Model 3	$N(0,1)$	0.25	0.155 (0.070)	0.133 (0.099)	0.423 (0.176)	0.153 (0.082)
		0.50	0.128 (0.068)	0.097 (0.048)	0.191 (0.089)	0.161 (0.070)
		0.75	0.151 (0.084)	0.111 (0.065)	0.434 (0.143)	0.179 (0.089)
	t_3	0.25	0.226 (0.178)	0.193 (0.112)	0.321 (0.151)	0.264 (0.143)
		0.50	0.137 (0.065)	0.111 (0.066)	0.159 (0.090)	0.170 (0.080)
		0.75	0.185 (0.091)	0.177 (0.121)	0.352 (0.136)	0.239 (0.150)
	$\chi^2(2)$	0.25	0.087 (0.064)	0.076 (0.045)	0.370 (0.161)	0.134 (0.072)
		0.50	0.186 (0.104)	0.174 (0.103)	0.237 (0.095)	0.235 (0.101)
		0.75	0.549 (0.317)	0.555 (0.472)	0.612 (0.228)	0.685 (0.382)

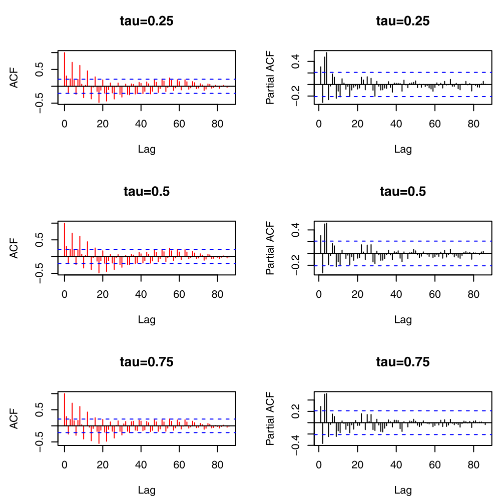

Figure 6. ACF and PACF plots of the residuals of model (12) at $\tau = 0.25$, 0.5, and 0.75.

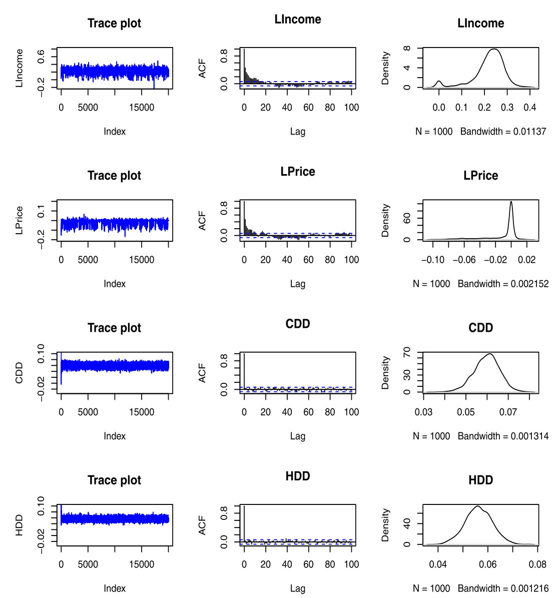

Figure 7. The first column displays one trace plot of 20,000 Gibbs samples of $BAL_{1/2}$ penalized estimation results of βs, the second and third columns represent the ACF plots and density plots of 1,000 sampled values of βs at $\tau = 0.5$ in the real example.

Table 8. $BL_{1/2}$ and $BAL_{1/2}$ results of electricity consumption data set

τ	Parameter	LIncome	LPrice	CDD	HDD	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	AQRRS
					Bayesian $L_{1/2}$ QR						
0.25	Est.	0.030	−0.051	0.059	0.057	0.520	−0.002	0.025	0.890	−0.463	0.008
	St.d	0.047	0.019	0.007	0.008	0.071	0.033	0.038	0.035	0.066	
	95% C.L.	−0.047	−0.090	0.045	0.041	0.366	−0.065	−0.053	0.819	−0.580	
	95% C.U.	0.137	−0.014	0.073	0.072	0.640	0.065	0.094	0.957	−0.315	
0.50	Est.	0.235	−0.043	0.060	0.058	0.621	−0.005	−0.000	0.911	−0.608	0.009
	St.d	0.057	0.027	0.007	0.005	0.090	0.027	0.031	0.035	0.091	
	95% C.L.	0.063	−0.089	0.046	0.048	0.457	−0.060	−0.060	0.845	−0.792	
	95% C.U.	0.315	0.007	0.072	0.068	0.814	0.053	0.062	0.980	−0.434	
0.75	Est.	0.457	−0.030	0.060	0.054	0.488	0.015	0.004	0.939	−0.477	0.008
	St.d	0.048	0.029	0.006	0.005	0.082	0.025	0.028	0.038	0.087	
	95% C.L.	0.370	−0.088	0.048	0.043	0.316	−0.035	−0.046	0.855	−0.625	
	95% C.U.	0.565	0.016	0.072	0.066	0.631	0.063	0.070	1.003	−0.293	
					Bayesian adaptive $L_{1/2}$ QR						
0.25	Est.	0.013	−0.046	0.058	0.058	0.539	0.001	0.015	0.894	−0.482	0.007
	St.d	0.037	0.023	0.007	0.008	0.062	0.026	0.032	0.037	0.060	
	95% C.L.	−0.025	−0.088	0.043	0.041	0.377	−0.052	−0.042	0.817	−0.578	
	95% C.U.	0.129	0.000	0.073	0.073	0.635	0.061	0.089	0.964	−0.332	
0.50	Est.	0.220	−0.011	0.060	0.056	0.684	−0.005	0.000	0.933	−0.684	0.009
	St.d	0.069	0.023	0.006	0.005	0.079	0.021	0.022	0.035	0.083	
	95% C.L.	0.000	−0.077	0.047	0.045	0.511	−0.038	−0.046	0.860	−0.834	
	95% C.U.	0.320	0.006	0.071	0.066	0.838	0.054	0.050	1.000	−0.484	
0.75	Est.	0.454	−0.004	0.059	0.053	0.523	0.011	−0.000	0.950	−0.521	0.008
	St.d	0.054	0.015	0.006	0.005	0.078	0.021	0.020	0.035	0.077	
	95% C.L.	0.357	−0.052	0.047	0.043	0.362	−0.029	−0.040	0.867	−0.652	
	95% C.U.	0.565	0.010	0.070	0.064	0.669	0.053	0.043	1.009	−0.351	

Received 12 September 2019

REFERENCES

[1] Alhamzawi, R. and Ali, H. T. M. (2018). Bayesian Tobit quantile regression with $L_{1/2}$ penalty. *Communications in Statistics-Simulation and Computation*, **47(6)**, 1739–1750. MR3236027

[2] Davino, C., Furno, M. and Vistocco, D. (2014). *Quantile Regression: Theory and Applications*. New York: John Wiley and Sons. MR1946581

[3] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96(456)**, 1348–1360. MR3230494

[4] Fallahpour, S. and Ahmed, S. E. (2014). Shrinkage estimation and variable selection in multiple regression models with random coefficient autoregressive errors. *Statistics and Probability Letters*, **92(5)**, 199–208. MR3334674

[5] Feng, X. N., Wang, G. C., Wang, Y. F. and Song, X. Y. (2015). Structure detection of semiparametric structural equation models with Bayesian adaptive group lasso. *Statistics in Medicine*, **34**, 1527–1547. MR3588567

[6] Feng, X. N., Wang, Y. F., Lu, B. and Song, X. Y. (2017). Bayesian regularized quantile structural equation models. *Journal of Multivariate Analysis*, **154**, 234–248. MR1646710

[7] Fu, W. J. (1998). Penalized regresssion: the bridge versus lasso. *Journal of Computational and Graphical Statistics*, **7(3)**, 397–416.

[8] Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8(1)**, 140–154. MR3580932

[9] Huang, Y. and Chen, J. (2016). Bayesian quantile regression-based nonlinear mixed-effects joint models for time-to-event and longitudinal data with multiple features. *Statistics in Medicine*, **35(30)**, 56–66. MR0611894

[10] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12(1)**, 55–67. MR3272559

[11] Jiang, Y. and Li, H. (2014). Penalized weighted composite quantile regression in the linear regression model with heavy-tailed autocorrelated errors. *Journal of the Korean Statistical Society*, **43(4)**, 531–543. MR3934810

[12] Kang, K.. Song, X. Y., Hu, J. and Zhu, H. T. (2019). Bayesian adaptive group lasso with semiparametric hidden Markov models. *Statistics in Medicine*, **38**, 1634–1650. MR1805787

[13] Knight, K. and Fu, W. J. (2000). Asymptotics for LASSO-type estimators. *Annals of Statistics*, **28(5)**, 1356–1378. MR0474644

[14] Koenker, R. and Bassett, J. (1978). Regression quantiles. *Econometrica*, **46(1)**, 33–50. MR2268657

[15] Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press. MR2851270

[16] Kozumi, H. and Kobayashi, G (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, **81(11)**, 1565–1578. MR2050474

[17] Lee, J. and Lund, R. (2004). Revisiting simple linear regression with autocorrelated errors. *Biometrika*, **91(1)**, 240–245.

[18] Liang, Y., liu, C. and Luan, X. Z., et al (2013). Sparse logistic regression with a $L_{1/2}$ penalty for gene selection in cancer classification. *BMC Bioinformatics*, **14(1)**, 198–198. MR3272558

[19] Lim, Y. and Oh, H. (2014). Variable selection in quantile regression when the models have autoregression errors. *Journal of Korean Statistical Society*, **43(4)**, 513–530. MR1711361

[20] Lin, T., Pourahmadi, M. and Schick, A. (2000). Regression models with time series errors. *Journal of Time Series Analysis*, **20(4)**, 425–433.

[21] Luan, X. Z., Liang, Y. and Liu, C., et al. (2014). A novel $L_{1/2}$ regularization shooting method for Cox's proportional hazards model. *Soft Computing*, **18(1)**, 143–152. MR3774527

[22] Mallick, H. and Yi, N. (2018). Bayesian bridge regression. *Journal of Applied Statistics*, **45(6)**, 988–1008. MR2524001

[23] Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, **103(482)**, 681–686. MR3248673

[24] Polson, N. G., Scott, J. G. and Windle, J. (2014). The Bayesian bridge. *Journal of the Royal Statistical Society: Series B*, **76(4)**, 713–733.

[25] Ramanathan, R. (2002). *Introductory Econometrics with Applications (5th Edition)*, Singapore: Singapore Thomson Asia.

[26] Reed, C. and Yu, K. (2009). *Efficient Gibbs sampling for Bayesian quantile regression*, Technical report, Brunel University, UK. MR2816698

[27] Reich, B. J., Fuentes, M. and Dunson, D. B. (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association*, **106(493)**, 6–20. MR3266041

[28] Rosadi, D. and Peiris, S. (2015). Second-order least-squares estimation for regression models with autocorrelated errors. *Computational Statistics*, **29(5)**, 931–943. MR0742860

[29] Tasy, R. S. (1984). Regression models with time series errors. *Journal of American Statistical Association*, **79(385)**, 118–124. MR3528645

[30] Tian, Y. Z, Li, E. Q. and Tian, M. Z. (2016). Bayesian joint quantile regression for mixed effects models with censoring and errors in covariates. *Computational Statistics*, **31(3)**, 1–27. MR3938055

[31] Tian, Y. Z, Shen, S. L., Lu, G., Tang, M. L. and Tian, M. Z. (2019). Bayesian LASSO-regularized quantile regression for linear regression models with autoregressive errors. *Communications in Statistics – Simulation and Computation*, **48(3)**, 777–796. MR2815776

[32] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **73(1)**, 273–282. MR2301500

[33] Wang, H., Li, G. and Tasy, C. (2007). Regression coefficients and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, **69(1)**, 63–78. MR3465645

[34] Wang, Y. F., Feng, X. N. and Song, X. Y. (2015). Bayesian quantile structural equation models. *Structural Equation Modeling – A Multidisciplinary Journal*, **23**, 246–258. MR3999240

[35] Wang, C. J., Li, Q., Song, X. Y. and Dong, X. G. (2019). Bayesian adaptive lasso for additive hazard regression with current status data. *Statistics in Medicine*, **38** 3703–3718. MR2903418

[36] Wu, R. and Wang, Q. (2012). Shrinkage estimation for linear regression with ARMA errors. *Journal of Statistical Planning and Inference*, **142(7)**, 2136–2148. MR2671451

[37] Xu, Z., Zhang, H., Wang, Y., Chang, X. and Liang, Y. (2010). $L_{1/2}$ regularization. *Science China Information Sciences*, **53(6)**, 1159–1169. MR2976867

[38] Yang, C. Y. (2012). Estimation of linear regression models with serially correlated errors. *Journal of Data Science*, **10(4)**, 723–755. MR3656115

[39] Yoon, Y. J., Lee, S. and Lee, T. (2017). Adaptive LASSO for linear regression models with ARMA-GARCH errors. *Communications in Statistics-Simulation and Computation*, **46(5)**, 3479–3490. MR1861390

[40] Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, **54(4)**, 437–447. MR3066690

[41] Yu, K., Chen, W. S., Reed, C., Dunson, D. B. (2013). Bayesian variable selection in quantile regression. *Statistics in Its Interface*, **6**, 261–274. MR3226770

[42] Zhang, H., Xu, Z. B., Wang, Y., Chang X. Y., Liang Y. (2014). A sharp nonasymptotic bound and phase diagram of $L_{1/2}$ regularization. *ACTA Mathematica Sinica-English Series*, **30(7)**, 1242–1258. MR3299347

[43] Zhao, K. F., Lian, H. (2015). Bayesian Tobit quantile regression with single-index models. *Journal of Statistical Computation and Simulation*, **85(6)**, 1247–1263. MR2279469

[44] Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, **101(476)**, 1418–1429. MR2137327

[45] Zou, H., Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, **167(2)**, 301–320.

Yuzhu Tian
School of Mathematics and Statistics
Northwest Normal University
LanZhou
PR China
E-mail address: pole1999@163.com

Xinyuan Song
Department of Statistics
The Chinese University of Hong Kong
Shatin, N.T.
Hong Kong
E-mail address: xysong@sta.cuhk.edu.hk