Bayesian Analysis (2020) **15**, Number 2, pp. 579–604

Bayesian Quantile Regression with Mixed Discrete and Nonignorable Missing Covariates

Zhi-Qiang Wang* and Nian-Sheng Tang†

Abstract. Bayesian inference on quantile regression (QR) model with mixed discrete and non-ignorable missing covariates is conducted by reformulating QR model as a hierarchical structure model. A probit regression model is adopted to specify missing covariate mechanism. A hybrid algorithm combining the Gibbs sampler and the Metropolis-Hastings algorithm is developed to simultaneously produce Bayesian estimates of unknown parameters and latent variables as well as their corresponding standard errors. Bayesian variable selection method is proposed to recognize significant covariates. A Bayesian local influence procedure is presented to assess the effect of minor perturbations to the data, priors and sampling distributions on posterior quantities of interest. Several simulation studies and an example are presented to illustrate the proposed methodologies.

MSC 2010 subject classifications: Primary 62F15, 62H12; secondary 62J20.

Keywords: Bayesian analysis, local influence analysis, non-ignorable missing data, quantile regression, variable selection.

1 Introduction

Quantile regression (QR) (Hendricks and Koenker, 1992; Hallock and Koenker, 2001; Chernozhukov, 2005; Gaglianone et al., 2011; Cade and Noon, 2003) has become an important tool for quantifying the conditional quantile relationship between a response variable and some covariates, due to its merits, such as, few assumptions on the distribution of random errors except for requiring that random errors have a zero conditional quantile, and more robustness to outliers and heavy-tailed data than ordinary least squares regression. QR analysis has received considerable attention over the past years. For example, see Koenker (2005) for a comprehensive overview, Geraci and Bottai (2006) for longitudinal data analysis, Wu and Liu (2009) for variable selection, Canay (2011) for panel data analysis. In particular, Bayesian analysis of QR models has been widely studied over the past years. For example, Yu and Moye (2001) studied Bayesian QR by reformulating QR as an asymmetric Laplace distribution; Reich et al. (2009) investigated Bayesian analysis of QR with independent and clustered data under the assumption that random errors are distributed as an infinite mixture of normals; Lancaster and Sung (2010) developed a Bayesian exponentially tilted empirical likelihood inference on QR; Yang and He (2012) presented a Bayesian empirical likelihood infer-

*Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, P. R. of China

†Correspondence to: Prof. Nian-Sheng Tang, Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, P. R. of China. Tel: 86-871-5032416, fax: 86-871-5033700, nstang@ynu.edu.cn

 https://doi.org/10.1214/19-BA1165

ence on QR; Kottas and Krnjajić (2009) proposed a Bayesian semiparametric method for QR using Dirichlet process mixtures to approximate the distribution of random errors; Huang and Chen (2016), Zhang et al. (2017) and Huang et al. (2017) discussed Bayesian QR-based mixed-effects joint models by formulating the asymmetric Laplace distribution as a mixture of normal and exponential distributions. The aforementioned works focus on the fully observed data.

However, missing data are commonly encountered in various fields such as clinical trials, social sciences and economics. QR analysis with missing data has widely been investigated in recent years. For example, Yi and He (2009), Wei et al. (2012), Wei and Yang (2014), Chen et al. (2015), Yuan and Yin (2010) and Huang (2016) studied the parameter estimation problem of QR models under the assumption that continuous covariates are subject to missingness at random (MAR). However, to our knowledge, there is little literature on variable selection and local influence analysis in Bayesian QR models with mixed discrete and continuous and nonignorable missing covariates though variable selection (Li et al., 2010; Alhamzawi et al., 2012) and local influence analysis (Cook, 1986; Zhu et al., 2011; Tang et al., 2017; Zhang and Tang, 2017; Ju et al., 2018) are two important steps in data analysis.

The main contribution of this paper includes that (i) a complicated QR model is considered by incorporating discrete and continuous and nonignorable missing covariates; (ii) a sequence of one-dimensional exponential family conditional distributions is adopted to specify the distribution of missing covariates because of discrete and continuous covariates involved; (iii) a sequence of one-dimensional probit regression models is employed to formulate nonignorable missingness covariates mechanism, which is easier to draw observations required for statistical inference from their corresponding conditional distributions than the widely used logistic regression models; (iv) a Bayesian adaptive LASSO procedure is developed to select covariates/explanatory variables in QR model and missingness covariates mechanism model; (v) Bayesian local influence analysis of Zhu et al. (2011) is extended to check the plausibility of missingness covariates mechanism in the considered QR model.

The rest of this paper is organized as follows. Section 2 introduces QR model with mixed discrete and continuous and nonignorable missing covariates. Section 3 investigates Bayesian inference. Section 4 discusses Bayesian variable selection. Bayesian local influence analysis is studied in Section 5. Simulation studies are conducted to investigate the finite sample performance of the proposed methods in Section 6. An example is illustrated in Section 7. Some concluding remarks are given in Section 8. Technical details are presented in the Supplementary Materials (Wang and Tang, 2019).

2 Model and notation

2.1 Quantile regression model

Consider the following quantile regression (QR) model

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where y_i is a response variable, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a $p \times 1$ vector of explanatory variables, which may be subject to missingness, $\boldsymbol{\beta}_\tau$ is a p-dimensional parameter vector to be estimated and τ is the quantile level ($0 < \tau < 1$), and ϵ_i is a random error. It is assumed that ϵ_i's are independently but not necessarily identically distributed, and $\Pr(\epsilon_i < 0|\boldsymbol{x}_i) = \tau$ for $i = 1, \ldots, n$. Thus, model (2.1) amounts to assuming $Q_{y_i}(\tau|\boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau$, where $Q_{y_i}(\tau|\boldsymbol{x}_i) = \inf\{y : F(y|\boldsymbol{x}_i) \geq \tau\}$ is the τth conditional quantile of y_i given $\boldsymbol{x}_i$, and $F(y|\boldsymbol{x}_i)$ is the conditional distribution of y_i given $\boldsymbol{x}_i$. When $\boldsymbol{x}_i$'s are completely observed, $\boldsymbol{\beta}_\tau$ can be estimated by minimizing the following objective function $\Omega(\boldsymbol{\beta}_\tau) = \sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau)$ over $\boldsymbol{\beta}_\tau$, where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the check function, and $I(\cdot)$ is an indicator function. Following Yu and Moye (2001), the aforementioned optimization problem can be formulated as the maximum likelihood estimation problem by assuming the asymmetric Laplace distribution (ALD) for ϵ_i with density

$$f_\tau(\epsilon_i) = \sigma\tau(1-\tau)\exp\{-\sigma\rho_\tau(\epsilon_i)\}, \tag{2.2}$$

where τ determines the skewness of distribution, and σ is a scale parameter. To wit, an estimator of $\boldsymbol{\beta}_\tau$ can be obtained by maximizing the following objective function: $\sigma^n\tau^n(1-\tau)^n \exp\{-\sigma\sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_\tau)\}$ over $\boldsymbol{\beta}_\tau$.

Following Kozumi and Kobayashi (2011), the asymmetric Laplace distribution (2.2) can be regarded as a mixture of an exponential distribution and a scaled normal distribution, i.e., $\epsilon_i = \kappa_1 v_i + \sqrt{\sigma^{-1}\kappa_2 v_i}u_i$, where $\kappa_1 = (1-2\tau)/\{\tau(1-\tau)\}$, $\kappa_2 = 2/\{\tau(1-\tau)\}$, v_i follows an exponential distribution with parameter σ^{-1} (i.e., $\text{Exp}(\sigma^{-1})$) whose density is $f(v_i|\sigma) = \sigma\exp(-\sigma v_i)$, and u_i follows the standard normal distribution (i.e., $N(0,1)$). Then, the QR model (2.1) can be written as the following hierarchical model

$$\begin{cases} y_i = \boldsymbol{x}_i^\top\boldsymbol{\beta}_\tau + \kappa_1 v_i + \sqrt{\kappa_2 v_i/\sigma}u_i, \\ v_i|\sigma \sim \text{Exp}\left(\sigma^{-1}\right) = \sigma\exp(-\sigma v_i), \\ u_i \sim N(0,1) = \dfrac{1}{\sqrt{2\pi}}\exp\left(-\dfrac{u_i^2}{2}\right). \end{cases} \tag{2.3}$$

When $\boldsymbol{x}_i$'s are subject to missingness but y_i's are completely observed, without loss of generality, it is assumed that $\boldsymbol{x}_{i,\text{mis}} = (x_{i1}, \ldots, x_{is_i})^\top$ and $\boldsymbol{x}_{i,\text{obs}} = (x_{i,s_i+1}, \ldots, x_{ip})^\top$, where s_i may vary across individuals. Thus, $\boldsymbol{x}_i$ can be written as $\boldsymbol{x}_i = \{\boldsymbol{x}_{i,\text{mis}}, \boldsymbol{x}_{i,\text{obs}}\}$. Let r_{ij} be missing indicator for x_{ij}, i.e., $r_{ij} = 0$ if x_{ij} is missing, and $r_{ij} = 1$ if x_{ij} is observed for $j = 1, \ldots, p$. Denote $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{ip})^\top$. Thus, the complete data set consists of observations $\{(y_i, \boldsymbol{x}_i, \boldsymbol{r}_i) : i = 1, \ldots, n\}$. Under the above assumptions, the joint probability density of responses, covariates and missing data indicators is given by

$$\begin{aligned} f(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{r}|\boldsymbol{\theta}) &= f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}_\tau, \sigma) f(\boldsymbol{x}_{\text{mis}}|\boldsymbol{\alpha}) f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\gamma}) \\ &= \prod_{i=1}^n f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}_\tau, \sigma) f(\boldsymbol{x}_{i,\text{mis}}|\boldsymbol{\alpha}) f(\boldsymbol{r}_i|y_i, \boldsymbol{x}_i, \boldsymbol{\gamma}), \end{aligned} \tag{2.4}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, $\boldsymbol{r} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n\}$, $\boldsymbol{x}_{\text{mis}} = \{\boldsymbol{x}_{1,\text{mis}}, \ldots, \boldsymbol{x}_{n,\text{mis}}\}$, $f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}_\tau, \sigma)$ is the marginal probability density function of y_i given $\boldsymbol{x}_i$ with unknown parameters $\boldsymbol{\beta}_\tau$ and σ, i.e., $f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}_\tau, \sigma) = \int f(y_i|\boldsymbol{x}_i, v_i, \boldsymbol{\beta}_\tau, \sigma) f(v_i|\sigma) dv_i$,

$f(\boldsymbol{x}_{i,\text{mis}}|\boldsymbol{\alpha})$ is the density function of missing covariates $\boldsymbol{x}_{i,\text{mis}}$ with an unknown parameter vector $\boldsymbol{\alpha}$, $f(\boldsymbol{r}_i|y_i,\boldsymbol{x}_i,\boldsymbol{\gamma})$ is the conditional function of $\boldsymbol{r}_i$ given y_i and $\boldsymbol{x}_i$ with an unknown parameter vector $\boldsymbol{\gamma}$, and $\boldsymbol{\theta}=\{\boldsymbol{\beta}_\tau,\sigma,\boldsymbol{\alpha},\boldsymbol{\gamma}\}$ contains all unknown distinct parameters. In the presence of nonignorable missing covariates, $f(\boldsymbol{r}|\boldsymbol{y},\boldsymbol{x},\boldsymbol{\gamma})$ is required in the complete-data joint probability density function. Our main interest aims at the posterior inference on $\boldsymbol{\theta}$ based on missing data indicator set $\boldsymbol{r}$ and the observed data $\boldsymbol{D}_{\text{obs}}=\{\boldsymbol{y},\boldsymbol{x}_{\text{obs}}\}$ in which $\boldsymbol{x}_{\text{obs}}=\{\boldsymbol{x}_{1,\text{obs}},\ldots,\boldsymbol{x}_{n,\text{obs}}\}$. According to the definition of the model considered above, the joint posterior density of $\boldsymbol{\theta}$ given $\boldsymbol{r}$ and $\boldsymbol{D}_{\text{obs}}$ is given by

$$f(\boldsymbol{\theta}|\boldsymbol{r},\boldsymbol{D}_{\text{obs}})\propto\left\{\prod_{i=1}^{n}\int_{\boldsymbol{x}_{i,\text{mis}}}f(y_i|\boldsymbol{x}_i,\boldsymbol{\beta}_\tau,\sigma)f(\boldsymbol{x}_{i,\text{mis}}|\boldsymbol{\alpha})f(\boldsymbol{r}_i|y_i,\boldsymbol{x}_i,\boldsymbol{\gamma})d\boldsymbol{x}_{i,\text{mis}}\right\}f(\boldsymbol{\theta}),\tag{2.5}$$

where $f(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$. Generally, the integral in Equation (2.5) does not have a closed form.

2.2 The distribution of missing covariates

Following Ibrahim et al. (1999), we can use a sequence of one-dimensional conditional distributions to specify the joint density $f(\boldsymbol{x}_{i,\text{mis}}|\boldsymbol{\alpha})$ of $\boldsymbol{x}_{i,\text{mis}}$ including discrete and continuous covariates, which is given by

$$f(\boldsymbol{x}_{i,\text{mis}}|\boldsymbol{\alpha})=f(x_{is_i}|x_{i1},\ldots,x_{i,s_i-1},\boldsymbol{\alpha}_{s_i})\cdots f(x_{i2}|x_{i1},\boldsymbol{\alpha}_2)f(x_{i1}|\boldsymbol{\alpha}_1),\tag{2.6}$$

where $\boldsymbol{\alpha}_k$ is an unknown parameter vector associated with the kth conditional distribution $f(x_{ik}|x_{i1},\ldots,x_{i,k-1},\boldsymbol{\alpha}_k)$ and the $\boldsymbol{\alpha}_k$'s are distinct for $k=1,\ldots,s_i$, and $\boldsymbol{\alpha}=\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_p\}$. Model (2.6) implies that it is easy to directly specify the distribution of missing covariates regardless of the discrete and continuous covariates. For example, one can specify the distribution of missing covariate x_{ik} via an exponential family distribution with the form

$$f(x_{ik}|x_{i1},\ldots,x_{i,k-1},\boldsymbol{\alpha}_k)=\exp\left\{\frac{x_{ik}\vartheta_{ik}-b_k(\vartheta_{ik})}{a_k(\phi_k)}-h_k(x_{ik},\phi_k)\right\},\tag{2.7}$$

where $a_k(\cdot)>0$, $b_k(\cdot)$ and $h_k(\cdot)$ are some appropriate known functions, ϑ_{ik} is the canonical parameter, ϕ_k is the dispersion parameter which is known or can be estimated separately, and $\mu_x^{ik}=E(x_{ik}|\cdot)=\dot{b}_k(\vartheta_{ik})$ and $\sigma_x^{ik}=\text{var}(x_{ik}|\cdot)=a_k(\phi_k)\ddot{b}_k(\vartheta_{ik})$ in which $\dot{b}_k(\vartheta)$ and $\ddot{b}_k(\vartheta)$ represent the first- and second-order derivatives of $b_k(\vartheta)$ with respect to ϑ, respectively. The density function of x_{ik} defined in Equation (2.7) includes the normal distribution, Bernoulli distribution, binomial/multinomial distribution, exponential distribution, Poisson distribution and gamma distribution as its special cases.

The relationship between x_{ik} and $\{x_{i1},\ldots,x_{i,k-1}\}$ can be formulated by

$$\eta_{ik}=g(\mu_x^{ik})=\alpha_{k0}+\alpha_{k1}x_{i1}+\ldots+\alpha_{k,k-1}x_{i,k-1},\tag{2.8}$$

where $g(\cdot)$, called the link function, is a known strictly monotone differentiable function, and $\boldsymbol{\alpha}_k=(\alpha_{k0},\alpha_{k1},\ldots,\alpha_{k,k-1})^\top$. Although we include covariates $\{x_{i1},\ldots,x_{i,k-1}\}$ in

specifying the conditional distribution of x_{ik}, only a few covariates may indeed have an effect on x_{ik}. Thus, variable selection procedure should be developed to select significant covariates. The model defined in Equations (2.7) and (2.8) is typically referred to as a generalized linear model (GLM), which includes normal linear regression model and logistic regression model as its special cases.

Note that model (2.6) requires to be specified only for those covariates that are subject to missingness. Also, one can specify a joint distribution for missing covariate vector $\boldsymbol{x}_{i,\text{mis}}$ via other approaches, for example, multivariate normal distribution for continuous missing covariates with p small (e.g., $\boldsymbol{x}_{i,\text{mis}} \sim N(\boldsymbol{\mu}_{xi}, \boldsymbol{\Sigma}_{xi})$), and multinomial distribution for discrete covariates with more than two possible values.

2.3 Models for missingness data mechanism

When covariates are subject to nonignorable missingness, we require setting up a missingness data mechanism model for Bayesian inference on $\boldsymbol{\theta}$ based on the observed data $\boldsymbol{D}_{\text{obs}}$. When r_{ij} is independent of r_{ik} for $j \neq k$, one possible model for $f(\boldsymbol{r}_i|\boldsymbol{x}_i, \boldsymbol{\gamma})$ is

$$\begin{aligned} f(\boldsymbol{r}_i|\boldsymbol{x}_i, \boldsymbol{\gamma}) &= \prod_{j=1}^{p} f(r_{ij}|\boldsymbol{x}_i, \boldsymbol{\gamma}) \\ &= \prod_{j=1}^{p} \{\Pr(r_{ij} = 1|\boldsymbol{x}_i, \boldsymbol{\gamma})\}^{r_{ij}} \{1 - \Pr(r_{ij} = 1|\boldsymbol{x}_i, \boldsymbol{\gamma})\}^{1-r_{ij}}. \end{aligned} \tag{2.9}$$

Following Ibrahim et al. (1999), one can relax the above assumption by using a sequence of one-dimensional conditional distributions to specify missingness data mechanism. To wit, one can write the conditional probability density function $f(\boldsymbol{r}_i|\boldsymbol{x}_i, \boldsymbol{\gamma})$ as

$$\begin{aligned} f(\boldsymbol{r}_i|\boldsymbol{x}_i, \boldsymbol{\gamma}) &= f(r_{ip}|\boldsymbol{r}_{i(p)}, \boldsymbol{x}_{i(p)}, \boldsymbol{\gamma}_p) f(r_{i,p-1}|\boldsymbol{r}_{i(p-1)}, \boldsymbol{x}_{i(p-1)}, \boldsymbol{\gamma}_{p-1}) \\ &\quad \times \ldots \times f(r_{i2}|r_{i1}, \boldsymbol{x}_{i(2)}, \boldsymbol{\gamma}_2) f(r_{i1}|x_{i1}, \boldsymbol{\gamma}_1), \end{aligned} \tag{2.10}$$

where $\boldsymbol{\gamma}_j$ is an unknown parameter vector associated with the conditional distribution of r_{ij} given $\{\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}\}$, $\boldsymbol{r}_{i(j)} = \{r_{i1}, \ldots, r_{i,j-1}\}$ and $\boldsymbol{x}_{i(j)} = \{x_{i1}, \ldots, x_{ij}\}$ for $j = 1, \ldots, p$, and $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p\}$. Since r_{ij} is binary variable, thus $f(r_{ij}|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j) = \{\Pr(r_{ij} = 1|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j)\}^{r_{ij}} \{1 - \Pr(r_{ij} = 1|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j)\}^{1-r_{ij}}$. Similar to Lee and Tang (2006), $\Pr(r_{ij} = 1|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j)$ can be formulated by $\text{logit}\{\Pr(r_{ij} = 1|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j)\} = \boldsymbol{x}_{zij}^{\top} \boldsymbol{\gamma}_j$, where $\text{logit}(a) = \log\{a/(1-a)\}$, $\boldsymbol{\gamma}_j = (\gamma_{j0}, \ldots, \gamma_{j,2j-1})^{\top}$, $\boldsymbol{x}_{zij} = (1, x_{i1}, \ldots, x_{ij}, r_{i1}, \ldots, r_{i,j-1})^{\top}$. In this case, it is rather difficult to draw observations from the conditional distribution $f(\boldsymbol{\gamma}_j|\boldsymbol{x}_{i(j)}, \boldsymbol{r}_{i(j)})$ in that $f(\boldsymbol{\gamma}_j|\boldsymbol{x}_{i(j)}, \boldsymbol{r}_{i(j)})$ is an unfamiliar distribution.

To address the issue, $\Pr(r_{ij} = 1|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j)$ is here formulated by the following probit regression model:

$$\Phi^{-1}\{\Pr(r_{ij} = 1|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j)\} = \boldsymbol{x}_{zij}^{\top} \boldsymbol{\gamma}_j, \tag{2.11}$$

where $\Phi^{-1}(\cdot)$ is the inverse function of the cumulative distribution function of the standard normal distribution. Thus, the probit model (2.11) can be reformulated as an underlying normal regression structure by introducing latent variables (Albert and Chib, 1993). To wit, introducing latent variables z_{ij} $(i = 1, \ldots, n, j = 1, \ldots, p)$, model (2.11)

can be rewritten as

$$\begin{aligned} r_{ij} &= \begin{cases} 1, & \text{if } z_{ij} > 0, \\ 0, & \text{if } z_{ij} \le 0, \end{cases} \\ z_{ij} &= \mu_{ij} + u_{ij}, \quad \mu_{ij} = \boldsymbol{x}_{zij}^{\top}\boldsymbol{\gamma}_j, \quad u_{ij} \sim N(0,1). \end{aligned} \tag{2.12}$$

Thus, conditional probability density function of r_{ij} given $\{\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}\}$ has the form of $f(r_{ij}|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j) = \int f(r_{ij}, z_{ij}|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j)dz_{ij} = \int f(z_{ij}|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j)\{I(r_{ij} = 1)I(z_{ij} > 0) + I(r_{ij} = 0)I(z_{ij} \le 0)\}dz_{ij}$, where $f(z_{ij}|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)}, \boldsymbol{\gamma}_j) \sim N(\boldsymbol{x}_{zij}^{\top}\boldsymbol{\gamma}_j, 1)$.

It is easily seen from the above argument that the probit model per se doesn't make a lot of difference with the logistic model, at least locally, the advantage of working with model (2.12) in comparison with logistic regression model is that the former is easier to draw observations required for Bayesian inference on $\boldsymbol{\beta}_\tau$ via the Gibbs sampler than the latter.

Note that we accommodate almost all the covariates $\{x_{i1}, \ldots, x_{i,j-1}, r_{i1}, \ldots, r_{i,j-1}\}$ in specifying missingness data mechanism of covariate x_{ij}, but only a few covariates may indeed contribute to missingness of covariate x_{ij}. To wit, the considered probit regression model has a sparse structure. In this case, variable selection procedure should be presented to distinguish important from unimportant covariates in the above considered probit/logistic regression model.

2.4 Prior distributions

To make Bayesian inference on $\boldsymbol{\theta}$ based on the observed data $\boldsymbol{D}_{\text{obs}}$ and missing data indicator set $\boldsymbol{r}$, it is necessary to first specify their prior distributions. Similar to Tang and Zhao (2014), we consider the following prior distributions for $\boldsymbol{\beta}_\tau$, σ, $\boldsymbol{\alpha}_k$ $(k = 1, \ldots, s_i)$ and $\boldsymbol{\gamma}_j$ $(j = 1, \ldots, p)$:

$$\boldsymbol{\beta}_\tau \sim N(\boldsymbol{\beta}_\tau^0, \boldsymbol{\Sigma}_{\tau\beta}^0), \quad \sigma \sim \Gamma(\alpha_\sigma^0, \beta_\sigma^0), \quad \boldsymbol{\alpha}_k \sim N(\boldsymbol{\alpha}_k^0, \boldsymbol{\Sigma}_{\alpha k}^0), \quad \boldsymbol{\gamma}_j \sim N(\boldsymbol{\gamma}_j^0, \boldsymbol{\Sigma}_{\gamma j}^0), \tag{2.13}$$

where $\boldsymbol{\beta}_\tau^0$, $\boldsymbol{\Sigma}_{\tau\beta}^0$, α_σ^0, β_σ^0, $\boldsymbol{\alpha}_k^0$, $\boldsymbol{\Sigma}_{\alpha k}^0$, $\boldsymbol{\gamma}_j^0$ and $\boldsymbol{\Sigma}_{\gamma j}^0$ are the hyperparameters whose values are assumed to be given by users or prior information, and $\Gamma(a_1, a_2)$ denotes the gamma distribution with parameters a_1 and a_2. The associated hyperparameters can be determined by a data-dependent prior based on auxiliary estimates of parameters. To wit, one can first evaluate Bayesian estimates of parameters with non-informative priors based on the training dataset, and then take the resultant Bayesian estimates as their corresponding hyperparameters in computing Bayesian estimates of parameters based on the testing dataset. Other kinds of data-dependent priors have been proposed in Bayesian analysis (Richardson and Green, 1997).

3 Conditional distributions and Bayesian estimation

It is easily seen from Equation (2.5) that it is rather difficult to make Bayesian inference on $\boldsymbol{\theta}$ via the posterior density function $f(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{D}_{\text{obs}})$ because of the intractable

high-dimensional integral involved. Here the data augmentation idea (Tanner and Wong, 1987) is adopted to address the issue. Following Tanner and Wong (1987), we augment missing data $\boldsymbol{x}_{\text{mis}}$ and latent variables $\{\boldsymbol{v}, \boldsymbol{z}\}$ with the observed data $\{\boldsymbol{r}, \boldsymbol{D}_{\text{obs}}\}$ in the posterior analysis, where $\boldsymbol{v} = \{v_1, \ldots, v_n\}$, and $\boldsymbol{z} = \{z_{ij} : i = 1, \cdots, n, j = 1, \cdots, p\}$. Thus, the posterior density $f(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{v}, \boldsymbol{z})$ is easier to handle than $f(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{D}_{\text{obs}})$ because the intractable multiple integral is not involved. But it is still rather difficult to sample observations $\boldsymbol{\theta}$ from the above presented posterior density because of some unfamiliar distributions involved. To this end, the Gibbs sampler is adopted to simulate a sequence of random observations from $f(\boldsymbol{x}_{\text{mis}}, \boldsymbol{v}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}_{\text{obs}}, \boldsymbol{r})$. In this algorithm, observations $\{\boldsymbol{x}_{\text{mis}}, \boldsymbol{v}, \boldsymbol{z}, \boldsymbol{\theta}\}$ are iteratively simulated from the following conditional distributions: $f(\boldsymbol{x}_{\text{mis}}|\boldsymbol{y}, \boldsymbol{x}_{\text{obs}}, \boldsymbol{r}, \boldsymbol{v}, \boldsymbol{z}, \boldsymbol{\theta})$, $f(\boldsymbol{v}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{\theta})$, $f(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{v}, \boldsymbol{\theta})$, and $f(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{v}, \boldsymbol{z})$. These conditional distributions are summarized in the Supplementary Materials. Convergence of the algorithm can be monitored by the estimated potential scale reduction (EPSR) values of parameters, which are computed sequentially as the runs proceed. If the EPSR values of all unknown parameters are less than 1.2, we claim that convergence of the Gibbs algorithm is attained (Gelman, 1996). Also, convergence can be investigated by inspecting several parallel sequences of observations drawn on the basis of different starting values. In particular, when the support of the distribution is nonconvex, the approach comparing between and within variances of multiple chains (Brooks and Andrew, 1998) can be used to monitor the convergence of the Gibbs algorithm.

Let $\{(\boldsymbol{\beta}_\tau^{(t)}, \sigma^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\alpha}^{(t)}) : t = 1, \ldots, \mathcal{T}\}$ be the observations of $\{\boldsymbol{\beta}_\tau, \sigma, \boldsymbol{\gamma}, \boldsymbol{\alpha}\}$ simulated from the joint conditional distribution $f(\boldsymbol{\beta}_\tau, \sigma, \boldsymbol{\gamma}, \boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{v}, \boldsymbol{z})$ via the above proposed Gibbs sampler. Bayesian estimates of parameters $\boldsymbol{\beta}_\tau, \sigma, \boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ can be obtained by

$$\hat{\boldsymbol{\beta}}_\tau = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \boldsymbol{\beta}_\tau^{(t)}, \quad \hat{\sigma} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \sigma^{(t)}, \quad \hat{\boldsymbol{\gamma}} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \boldsymbol{\gamma}^{(t)}, \quad \hat{\boldsymbol{\alpha}} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \boldsymbol{\alpha}^{(t)}.$$

Similarly, we can use the simulated observations $\{(\boldsymbol{\beta}_\tau^{(t)}, \sigma^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\alpha}^{(t)}) : t = 1, \ldots, \mathcal{T}\}$ to obtain consistent estimates of their corresponding posterior covariance matrices. For example, the posterior covariance matrix $\text{var}(\hat{\boldsymbol{\beta}}_\tau|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \sigma)$ can be consistently estimated by

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_\tau|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \sigma) = \frac{1}{\mathcal{T} - 1} \sum_{t=1}^{\mathcal{T}} (\boldsymbol{\beta}_\tau^{(t)} - \hat{\boldsymbol{\beta}}_\tau)(\boldsymbol{\beta}_\tau^{(t)} - \hat{\boldsymbol{\beta}}_\tau)^\top.$$

Thus, the standard errors of components of $\hat{\boldsymbol{\beta}}_\tau$ can be obtained by the diagonal elements of $\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_\tau|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \sigma)$.

4 Bayesian variable selection

Variable selection plays an important role in the statistical model-building process. In many applications, one usually includes a large number of potential explanatory variables in an initially posited model. However, it is undesirable to cover some unimportant

explanatory variables in the final model because only a small number of explanatory variables indeed contribute to response variable. A popular approach to select significant explanatory variables is to add a penalty term to the objective function. The widely used approaches include LASSO (Tibshirani, 1996) and adaptive LASSO (Zou, 2006). Due to some merits such as the oracle property of the adaptive LASSO, they have been extended to QR model in recent years. For example, see Koenker (2004), Wu and Liu (2009) and Alhamzawi et al. (2012). In particular, Park and Casella (2008) pointed out that Bayesian adaptive LASSO for QR model can be implemented by specifying a symmetric Laplace prior for each of regression parameters $\beta_{\tau j}$'s, where $\beta_{\tau j}$ is the jth component of $\boldsymbol{\beta}_\tau$.

Following Park and Casella (2008), if we assume a conditional Laplace prior for $\beta_{\tau j}$ $(j = 1, \ldots, p)$, i.e.,

$$f(\boldsymbol{\beta}_\tau|\zeta, \lambda) = \prod_{j=1}^{p} \frac{\zeta\lambda}{2} \exp(-\zeta\lambda|\beta_{\tau j}|), \tag{4.1}$$

thus QR model becomes a LASSO QR model, which has a regularization effect for simultaneous parameter estimation and variable selection. Andrews and Mallows (1974) noted that the Laplace prior can be represented as a scale mixture of a normal distribution with an exponential mixing density, i.e.,

$$\frac{a}{2}\exp\{-a|t|\} = \int_0^\infty \frac{1}{\sqrt{2\pi\nu}} \exp\left(-\frac{t^2}{2\nu}\right) \frac{a^2}{2} \exp\left(-\frac{a^2\nu}{2}\right) d\nu.$$

Let $b_\beta = \zeta\lambda$. Then, the above proposed Laplace prior on $\beta_{\tau j}$ can be rewritten as

$$f(\beta_{\tau j}|b_\beta) = \frac{b_\beta}{2}\exp\{-b_\beta|\beta_{\tau j}|\} = \int_0^\infty \frac{1}{\sqrt{2\pi\nu_j}} \exp\left(-\frac{\beta_{\tau j}^2}{2\nu_j}\right) \frac{b_\beta^2}{2} \exp\left(-\frac{b_\beta^2\nu_j}{2}\right) d\nu_j.$$

To wit, Bayesian adaptive LASSO QR shows that $\beta_{\tau j}$ takes a normal prior (i.e., $\beta_{\tau j}|\nu_j \sim N(0, \nu_j)$) and parameter ν_j has an exponential prior (i.e., $\nu_j \sim \text{Exp}(b_\beta^2/2)$). A small b_β corresponds to a nonzero parameter, while a big b_β corresponds to a zero parameter. Leng et al. (2014) pointed out that the adaptive LASSO tuning parameter b_β can be obtained by empirical Bayesian method or hierarchical Bayesian method. Here the latter is employed to select b_β. In particular, b_β^2 is assumed to follow the gamma distribution with parameters δ and ψ, i.e., $b_\beta^2 \sim \Gamma(\delta, \psi)$, where δ and ψ are the hyperparameters. Generally, δ and ψ can be fixed at some small values to get a flat prior, or one can first fix δ and then estimate ψ by empirical Bayesian method. For the latter, similar to Casella (2001), at the kth iteration, given the current value $\psi^{(k-1)}$, ψ can be updated by $\psi^{(k)} = p\delta/\sum_{l=1}^{p} E(b_\beta^2|\psi^{(k-1)}, \nu_l)$, where $E(\cdot)$ represents the expectation taken with respect to the posterior distribution of b_β given $\{\psi^{(k-1)}, \nu_l\}$. According to our experience, as parameter δ lies in the deeper level of Bayesian hierarchical model, its value has little effect on statistical inference. Hence, in our simulation studies, we take $\delta = 0.1$ to get a flat prior. Similarly, we can use the above method to specify the priors of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

Under the above notation, the priors of $\boldsymbol{\beta}_\tau$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ can be written as the following hierarchical structures:

$$\begin{aligned}
&\boldsymbol{\beta}_\tau \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\beta), \quad \boldsymbol{\Sigma}_\beta = \mathrm{diag}(\nu_1^\beta, \ldots, \nu_p^\beta), \quad \nu_j^\beta \sim \mathrm{Exp}(b_{\beta j}^2/2), \quad b_{\beta j}^2 \sim \Gamma(\delta_\beta, \psi_\beta),\\
&\boldsymbol{\alpha}_k^* \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\alpha^k), \ \boldsymbol{\Sigma}_\alpha^k = \mathrm{diag}(\nu_{k1}^\alpha, \ldots, \nu_{k,k-1}^\alpha), \ \nu_{k,k_1}^\alpha \sim \mathrm{Exp}(b_{\alpha k k_1}^2/2), \ b_{\alpha k k_1}^2 \sim \Gamma(\delta_\alpha, \psi_\alpha),\\
&\boldsymbol{\gamma}_j^* \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\gamma^j), \ \boldsymbol{\Sigma}_\gamma^j = \mathrm{diag}(\nu_{j1}^\gamma, \ldots, \nu_{j,2j-1}^\gamma), \ \nu_{j,k_2}^\gamma \sim \mathrm{Exp}(b_{\gamma j k_2}^2/2), \ b_{\gamma j k_2}^2 \sim \Gamma(\delta_\gamma, \psi_\gamma),
\end{aligned}$$

for $k_1 = 1, \ldots, k-1$ and $k_2 = 1, \ldots, 2j-1$, where δ_β, ψ_β, δ_α, ψ_α, δ_γ and ψ_γ are the hyperparameters whose values are pregiven by users, $\boldsymbol{\alpha}_k^* = (\alpha_{k1}, \ldots, \alpha_{k,k-1})^\top$ for $k = 1, \ldots, s_i$ and $\boldsymbol{\gamma}_j^* = (\gamma_{j1}, \ldots, \gamma_{j,2j-1})^\top$ for $j = 1, \ldots, p$.

For the above discussed variable selection procedure, the Gibbs sampler introduced in Section 3 can be employed to sample a series of random observations from the joint conditional distribution $f(\boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{v}, \boldsymbol{z}, \sigma, \boldsymbol{\beta}_\tau, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \boldsymbol{y}, \boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{r})$ based on the above specified priors of $\boldsymbol{\beta}_\tau$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

5 Bayesian local influence analysis

In this section, we develop a Bayesian local influence analysis approach to assess the effect of minor perturbations to the data, priors and sampling distributions on the posterior quantities of interest in QR.

5.1 Bayesian perturbation model and manifold

Similar to Zhu et al. (2011), we consider a class of perturbation models simultaneously perturbing to the data, priors and sampling distributions:

$$\begin{aligned}
f(\boldsymbol{y}, \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{v}, \boldsymbol{\theta} | \boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{\omega}) \quad = \quad & f(\boldsymbol{\theta}|\boldsymbol{\omega}_p) \prod_{i=1}^n \{ f(y_i, v_i | \boldsymbol{x}_i, \boldsymbol{\beta}, \sigma, \boldsymbol{\omega}_d, \boldsymbol{\omega}_s) \\
& f(\boldsymbol{x}_{i,\mathrm{mis}} | \boldsymbol{\alpha}, \boldsymbol{\omega}_d, \boldsymbol{\omega}_s) f(\boldsymbol{z}_i, \boldsymbol{r}_i | y_i, \boldsymbol{x}_i, \boldsymbol{\gamma}, \boldsymbol{\omega}_d, \boldsymbol{\omega}_s) \},
\end{aligned}$$

which satisfies $\int f(\boldsymbol{y}, \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{v}, \boldsymbol{\theta} | \boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{\omega}) d\boldsymbol{y} d\boldsymbol{z} d\boldsymbol{x}_{\mathrm{mis}} d\boldsymbol{v} d\boldsymbol{\theta} = 1$, where $\boldsymbol{\omega}_p \in \mathcal{R}^{m_p}$, $\boldsymbol{\omega}_d \in \mathcal{R}^{m_d}$ and $\boldsymbol{\omega}_s \in \mathcal{R}^{m_s}$ represent perturbations to the priors, the data and the sampling distributions, respectively. Let $m = m_p + m_d + m_s$. It is assumed that $\boldsymbol{\omega}^0 = (\boldsymbol{\omega}_p^0, \boldsymbol{\omega}_d^0, \boldsymbol{\omega}_s^0) \in \mathcal{R}^m$ represents no perturbation.

Following the argument of Zhu et al. (2011), under some regularity conditions, the perturbed model $\mathcal{M} = \{ f(\boldsymbol{y}, \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{v}, \boldsymbol{\theta} | \boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{\omega}) : \boldsymbol{\omega} \in \mathcal{R}^m \}$ can be regarded as a function in $\boldsymbol{\omega}$, and it forms an m-dimensional manifold. Then, the tangent space $T_{\boldsymbol{\omega}}$ of $\mathcal{M}$ at $\boldsymbol{\omega}^0 \in \mathcal{M}$ is spanned by m functions: $\partial_{\omega_k} \ell(\boldsymbol{\omega}^0) = \partial \ell(\boldsymbol{\omega}) / \partial \omega_k |_{\boldsymbol{\omega} = \boldsymbol{\omega}^0}$ for $k = 1, \ldots, m$, the m^2 quantities $g_{jk}(\boldsymbol{\omega}) = E_{\boldsymbol{\omega}} \{ \partial_{\omega_j} \ell(\boldsymbol{\omega}) \partial_{\omega_k} \ell(\boldsymbol{\omega}) \} = E_{\boldsymbol{\omega}} \{ -\partial_{\omega_j \omega_k}^2 \ell(\boldsymbol{\omega}) \}$ for $j, k = 1, \ldots, m$ form a metric tensor of $\mathcal{M}$, where $\ell(\boldsymbol{\omega}) = \ln f(\boldsymbol{y}, \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{v}, \boldsymbol{\theta} | \boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{\omega})$, ω_k is the kth component of $\boldsymbol{\omega}$, and $E_{\boldsymbol{\omega}}$ denotes the expectation taken with respect to the joint probability density function $f(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{v}, \boldsymbol{\theta} | \boldsymbol{x}_{\mathrm{obs}})$ and $\partial_{\omega_j \omega_k}^2 = \partial / \partial \omega_j \partial \omega_k$. Denote $\boldsymbol{G}(\boldsymbol{\omega}) = (g_{jk}(\boldsymbol{\omega}))$. The jth diagonal element g_{jj} of $\boldsymbol{G}(\boldsymbol{\omega}^0)$ quantifies the amount of the jth perturbation introduced by ω_j, the quantity $\rho_{jk} = g_{jk}(\boldsymbol{\omega}) / \sqrt{g_{jj}(\boldsymbol{\omega}) g_{kk}(\boldsymbol{\omega})}$ measures

the amount of the association between ω_j and ω_k. If $\boldsymbol{G}(\boldsymbol{\omega}^0)$ is a diagonal matrix, thus all components of $\boldsymbol{\omega}$ are orthogonal to each other, and the corresponding perturbation scheme is referred to as an appropriate perturbation. When $\boldsymbol{G}(\boldsymbol{\omega}^0)$ is not a diagonal matrix, we can always select a new perturbation vector $\widetilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^0 + \boldsymbol{G}(\boldsymbol{\omega}^0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}^0)$ such that $\boldsymbol{G}(\widetilde{\boldsymbol{\omega}})$ evaluated at $\boldsymbol{\omega}^0$ equals $c\boldsymbol{I}_m$, where c is a positive scalar. We consider the following perturbation schemes.

Example 1. Consider the perturbation to the priors of $\boldsymbol{\beta}_\tau$, $\boldsymbol{\gamma}_j$, $\boldsymbol{\alpha}_k$ and σ by assuming $\boldsymbol{\beta}_\tau|\omega_\beta \sim N_p(\boldsymbol{\beta}_\tau^0, \omega_\beta^{-1}\boldsymbol{\Sigma}_{\tau\beta}^0)$, $\boldsymbol{\gamma}_j|\omega_\gamma \sim N(\boldsymbol{\gamma}_j^0, \omega_\gamma^{-1}\boldsymbol{\Sigma}_{\gamma j}^0)$, $\boldsymbol{\alpha}_k|\omega_\alpha \sim N(\boldsymbol{\alpha}_k^0, \omega_\alpha^{-1}\boldsymbol{\Sigma}_{\alpha k}^0)$, and $\sigma|\omega_\sigma \sim \Gamma(\alpha_\sigma^0, \omega_\sigma\beta_\sigma^0)$, where ω_β, ω_γ, ω_α and ω_σ are the positive scalars. In this case, $\boldsymbol{\omega}_p = (\omega_\beta, \omega_\gamma, \omega_\alpha, \omega_\sigma)^\top$, and $\boldsymbol{\omega}_p^0 = (\omega_\beta^0, \omega_\gamma^0, \omega_\alpha^0, \omega_\sigma^0)^\top = (1, 1, 1, 1)^\top$ represents no perturbation. The perturbed model $\mathcal{M} = \{f(\boldsymbol{y}, \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{x}_{\text{mis}}, \boldsymbol{v}, \boldsymbol{\theta}|\boldsymbol{x}_{\text{obs}}, \boldsymbol{\omega}_p) : \boldsymbol{\omega}_p \in R^4\}$ forms a Riemannian manifold. The tangent space T_{ω_p} of $\mathcal{M}$ at $\boldsymbol{\omega}_p^0$ is spanned by

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\omega}_p^0)}{\partial \boldsymbol{\omega}_p} \;=\; \Big(&\frac{p}{2} - \frac{1}{2}(\boldsymbol{\beta}_\tau - \boldsymbol{\beta}_\tau^0)^\top(\boldsymbol{\Sigma}_{\tau\beta}^0)^{-1}(\boldsymbol{\beta}_\tau - \boldsymbol{\beta}_\tau^0), \frac{p_{\gamma j}}{2} - \frac{1}{2}(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^0)^\top(\boldsymbol{\Sigma}_{\gamma j}^0)^{-1}(\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_j^0),\\ &\frac{p_{\alpha k}}{2} - \frac{1}{2}(\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_k^0)^\top(\boldsymbol{\Sigma}_{\alpha k}^0)^{-1}(\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_k^0), \alpha_\sigma^0 - \beta_\sigma^0\sigma\Big)^\top,\end{aligned}$$

where $p_{\gamma j} = 2j - 1$ and $p_{\alpha k} = k - 1$. It is easily shown that $\boldsymbol{G}(\boldsymbol{\omega}_p^0) = \text{diag}(p/2, p_{\gamma j}/2, p_{\alpha j}/2,\ \alpha_\sigma^0)$. This perturbation scheme can be used to evaluate the effect of slightly disturbed prior distributions of $\boldsymbol{\beta}_\tau$, $\boldsymbol{\gamma}_j$, $\boldsymbol{\alpha}_k$ and σ.

Example 2. For probit missingness data mechanism model: $\Pr(r_{ij} = 1|\boldsymbol{x}_i, \boldsymbol{\gamma}_j) = \Phi(\gamma_{j0} + \gamma_{j1}x_{i1} + \ldots + \gamma_{jj}x_{ij})$, we consider the following perturbation scheme:

$$f(r_{ij}|\boldsymbol{x}_i, \boldsymbol{\gamma}_j, \omega_{\gamma j}) = \{\Phi(\boldsymbol{\chi}_{\gamma ij}^\omega)\}^{r_{ij}}\{1 - \Phi(\boldsymbol{\chi}_{\gamma ij}^\omega)\}^{1-r_{ij}},$$

where $\boldsymbol{\chi}_{\gamma ij}^\omega = \gamma_{j0} + \gamma_{j1}x_{i1} + \cdots + \gamma_{j,j-1}x_{i,j-1} + \omega_{\gamma j}x_{ij}$. In this scheme, $\omega_{\gamma_j}^0 = 0$ represents no perturbation. The tangent space $T_{\omega_{\gamma j}}$ of $\mathcal{M}$ at $\omega_{\gamma j}^0$ is spanned by

$$\frac{\partial \ell(\omega_{\gamma j}^0)}{\partial \omega_{\gamma j}} = \sum_{i=1}^n \Big[\Big\{\frac{r_{ij}}{\Phi(\varpi_{ij})} - \frac{1 - r_{ij}}{1 - \Phi(\varpi_{ij})}\Big\}\varphi(\varpi_{ij})x_{ij}\Big],$$

where $\varpi_{ij} = \gamma_{j0} + \gamma_{j1}x_{i1} + \ldots + \gamma_{j,j-1}x_{i,j-1}$, and $\varphi(\cdot)$ is the probability density function of the standard normal distribution. It is easily shown that $G(\omega_{\gamma j}^0)$ has the form

$$G(\omega_{\gamma j}^0) = E_{x_{ij}^*, \gamma_j^*, \alpha_j}\left\{\Big(a_j(\phi_j)\ddot{b}_j(\vartheta_{ij}) + \{\dot{b}_j(\vartheta_{ij})\}^2\Big)\sum_{i=1}^n \frac{\varphi^2(\varpi_{ij})}{\Phi(\varpi_{ij})(1 - \Phi(\varpi_{ij}))}\right\},$$

where $g(\dot{b}(\vartheta_{ij})) = \alpha_{j0} + \alpha_{j1}x_{i1} + \ldots + \alpha_{j,j-1}x_{i,j-1}$, $E_{x_{ij}^*, \gamma_j^*, \alpha_j}$ represents the expectation taken with respect to the joint distribution of $(\boldsymbol{x}_{ij}^*, \boldsymbol{\gamma}_j^*, \boldsymbol{\alpha}_j)$ in which $\boldsymbol{x}_{ij}^* = (x_{i1}, \ldots, x_{i,j-1})^\top$ and $\boldsymbol{\gamma}_j^* = (\gamma_{j0}, \gamma_{j1}, \ldots, \gamma_{j,j-1})^\top$. This perturbation scheme is used to measure the effect of tinily disturbed missingness data mechanism.

5.2 Bayesian local influence measures

Let $\boldsymbol{f}(\boldsymbol{\omega}) : \mathcal{M} \to \mathcal{R}^l$ be a l-dimensional objective function vector such as ϕ-divergence, Bayes factor and posterior mean distance of parameters of interest. For the finite dimensional manifold $\mathcal{M}$, if $\boldsymbol{\omega}(t)$ is a geodesic on $\mathcal{M}$ with $\boldsymbol{\omega}(0) = \boldsymbol{\omega}^0$ and $\partial_t \boldsymbol{\omega}(t)|_{t=0} = \boldsymbol{h} \in \mathcal{R}^m$, thus it follows from Taylor expansion that $\boldsymbol{f}(\boldsymbol{\omega}(t)) = \boldsymbol{f}(\boldsymbol{\omega}(0)) + \boldsymbol{f}'_h(0)t + O(t^2)$, where $\boldsymbol{f}'_h(0) = \boldsymbol{\nabla}_f^\top \boldsymbol{h}$ in which $\boldsymbol{\nabla}_f = \partial_\omega \boldsymbol{f}(\boldsymbol{\omega}(0))$.

If $\boldsymbol{\nabla}_f \neq 0$, the first-order influence (FI) measure at $\boldsymbol{\omega}^0$ in the direction vector $\boldsymbol{h} \in \mathcal{R}^m$ of unit length (i.e., the Euclidean norm of vector $\boldsymbol{h}$ is 1) is defined as

$$FI_{\boldsymbol{f},\boldsymbol{h}} = FI_{\boldsymbol{f}(\boldsymbol{\omega}(0)),\boldsymbol{h}} = \frac{\boldsymbol{h}^\top \boldsymbol{\nabla}_f \boldsymbol{W}_f \boldsymbol{\nabla}_f^\top \boldsymbol{h}}{\boldsymbol{h}^\top \boldsymbol{G} \boldsymbol{h}},$$

where $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\omega}^0)$ and $\boldsymbol{W}_f$ is some user-specified positive semi-definite matrix. In particular, for an appropriate perturbation $\widetilde{\boldsymbol{\omega}}$, $FI_{\boldsymbol{f},\boldsymbol{h}}$ can be rewritten as

$$FI_{\boldsymbol{f}(\widetilde{\omega}),\boldsymbol{h}|\widetilde{\omega}=\boldsymbol{\omega}^0} = \boldsymbol{h}^\top \boldsymbol{G}^{-1/2} \boldsymbol{\nabla}_f \boldsymbol{W}_f \boldsymbol{\nabla}_f^\top \boldsymbol{G}^{-1/2} \boldsymbol{h}.$$

Following Poon and Poon (1999), the first-order adjusted influence measure $FIC_{\boldsymbol{f}(\widetilde{\omega}^0),\boldsymbol{h}}$ at $\boldsymbol{\omega}^0$ in an unit direction vector $\boldsymbol{h}$ can be defined as

$$FIC_{\boldsymbol{f}(\widetilde{\omega}^0),\boldsymbol{h}} = \boldsymbol{h}^\top \mathcal{B} \boldsymbol{h},$$

where $\mathcal{B} = \boldsymbol{Q}/\mathrm{trace}(\boldsymbol{Q})$ in which $\boldsymbol{Q} = \boldsymbol{G}^{-1/2} \boldsymbol{\nabla}_f \boldsymbol{W}_f \boldsymbol{\nabla}_f^\top \boldsymbol{G}^{-1/2}$.

Similar to Poon and Poon (1999) and Zhu et al. (2014), $M(0)_j = FIC_{\boldsymbol{f}(\widetilde{\omega}^0),\boldsymbol{e}_j} = b_{jj}$ for $j = 1, \ldots, m$ can be used to assess the effect of various minor perturbations, where $\boldsymbol{e}_j$ is a basic perturbation vector with the jth element 1 and 0 elsewhere, b_{jj} is the jth diagonal element of matrix $\boldsymbol{B}$. We can use $\overline{M}(0) + 2SM(0)$ as a benchmark, where $M(0)$ and $SM(0)$ are the mean and standard error of $\{M(0)_j : j = 1, \ldots, m\}$, respectively.

Example 3 (Bayes factor)**.** We take $\boldsymbol{f}(\boldsymbol{\omega})$ to be Bayes factor defined by $B_F(\boldsymbol{\omega}) = \ln f(\boldsymbol{D}_{\mathrm{obs}}, \boldsymbol{r}|\boldsymbol{\omega}) - \ln f(\boldsymbol{D}_{\mathrm{obs}}, \boldsymbol{r}|\boldsymbol{\omega}^0)$, where $f(\boldsymbol{D}_{\mathrm{obs}}, \boldsymbol{r}|\boldsymbol{\omega}) = \int f(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}, \boldsymbol{r}, \boldsymbol{\theta}|\boldsymbol{\omega}) d\boldsymbol{z} d\boldsymbol{v}\, d\boldsymbol{x}_{\mathrm{mis}} d\boldsymbol{\theta}$. In this case, it is easily shown that $\boldsymbol{\nabla}_B = E\{\partial_\omega \ln f(\boldsymbol{y}, \boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{z}, \boldsymbol{v}, \boldsymbol{r}, \boldsymbol{\theta}|\boldsymbol{\omega}^0)|\boldsymbol{D}_{\mathrm{obs}}, \boldsymbol{r}, \boldsymbol{\omega}^0\}$, where $E\{\cdot|\cdot\}$ represents the expectation taken with respect to the conditional distribution $f(\boldsymbol{z}, \boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{v}, \boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{r})$. In this case, $\boldsymbol{\nabla}_B$ can be approximated by $\boldsymbol{\nabla}_B \approx S_0^{-1} \sum_{s=1}^{S_0} \partial_\omega \ln f(\boldsymbol{y}, \boldsymbol{x}_{\mathrm{mis}}^{(s)}, \boldsymbol{z}^{(s)}, \boldsymbol{v}^{(s)}, \boldsymbol{r}, \boldsymbol{\theta}^{(s)}|\boldsymbol{\omega}^0)$, where $\{(\boldsymbol{x}_{\mathrm{mis}}^{(s)}, \boldsymbol{z}^{(s)}, \boldsymbol{v}^{(s)}, \boldsymbol{\theta}^{(s)}) : s = 1, \ldots, S_0\}$ are generated from the joint posterior distribution $f(\boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{z}, \boldsymbol{v}, \boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{r})$ via the above introduced Gibbs sampler.

When $\nabla_f = 0$, it follows from Taylor expansion $f(\boldsymbol{\omega}(t))$ at $t = 0$ that $f(\boldsymbol{\omega}(t)) = f(\boldsymbol{\omega}(0)) + 0.5 f''_h(0) t^2 + O(t^3)$, where $f''_h(0) = \boldsymbol{h}^\top \boldsymbol{H}_f \boldsymbol{h}$ with $\boldsymbol{H}_f = \partial^2 f(\boldsymbol{\omega}) / \partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top|_{\boldsymbol{\omega}=\boldsymbol{\omega}^0}$. Similar to Zhu et al. (2011), we define the second-order influence measure (SI) at $\boldsymbol{\omega}^0$ in the unit direction vector $\boldsymbol{h} \in \mathcal{R}^m$ as

$$SI_{f,\boldsymbol{h}} = SI_{f(\boldsymbol{\omega}(0)),\boldsymbol{h}} = \frac{\boldsymbol{h}^\top \boldsymbol{H}_f \boldsymbol{h}}{\boldsymbol{h}^\top \boldsymbol{G} \boldsymbol{h}}.$$

Thus, for an appropriate perturbation $\widetilde{\boldsymbol{\omega}}$, $SI_{f,\boldsymbol{h}}$ reduces to

$$SI_{f(\widetilde{\boldsymbol{\omega}}),\boldsymbol{h}|\widetilde{\boldsymbol{\omega}}=\boldsymbol{\omega}^0} = \boldsymbol{h}^\top \boldsymbol{G}^{-1/2}\boldsymbol{H}_f\boldsymbol{G}^{-1/2}\boldsymbol{h}.$$

Again, we define the second-order adjusted influence measure at $\boldsymbol{\omega}^0$ in the unit direction vector $\boldsymbol{h}$ as $SIC_{f(\widetilde{\boldsymbol{\omega}}^0),\boldsymbol{h}} = \boldsymbol{h}^\top\mathcal{B}_S\boldsymbol{h}$, where $\mathcal{B}_S = \boldsymbol{Q}_S/\text{trace}(\boldsymbol{Q}_S)$ in which $\boldsymbol{Q}_S = \boldsymbol{G}^{-1/2}\boldsymbol{H}_f\boldsymbol{G}^{-1/2}$. The diagonal elements of matrix $\mathcal{B}_S$ can be used to identify the potential influential observations, misspecified priors and inappropriate modeling assumptions.

Example 4 (ϱ-divergence)**.** We take the objective function $f(\boldsymbol{\omega})$ to be the ϱ-divergence between two posterior probability density functions before and after introducing perturbation $\boldsymbol{\omega}$, which is defined by

$$D_\varrho(\boldsymbol{\omega}) = \int \varrho(R(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r},\boldsymbol{\omega}))f(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r})d\boldsymbol{x}_{\text{mis}}d\boldsymbol{z}d\boldsymbol{v}d\boldsymbol{\theta},$$

where $R(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r},\boldsymbol{\omega}) = f(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r},\boldsymbol{\omega})/f(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r})$, and $\varrho(\cdot)$ is a convex function with $\varrho(1) = 0$ such as the Kullback-Leibler divergence or the χ^2-divergence. Thus, we have $\boldsymbol{\nabla}_\varrho = 0$ and

$$\boldsymbol{H}_\varrho = \ddot{\varrho}(1)\left[E_{\omega^0}\{\dot{\ell}_\varrho(\boldsymbol{\omega}^0)|\boldsymbol{D}_{\text{obs}},\boldsymbol{r}\}^{\otimes 2} - \{E_{\omega^0}(\dot{\ell}_\varrho(\boldsymbol{\omega}^0)|\boldsymbol{D}_{\text{obs}},\boldsymbol{r})\}^{\otimes 2}\right],$$

where $\dot{\ell}_\varrho(\boldsymbol{\omega}^0) = \partial_\omega \log f(\boldsymbol{y},\boldsymbol{r},\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{x}_{\text{obs}},\boldsymbol{\omega})|_{\boldsymbol{\omega}=\boldsymbol{\omega}^0}$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^\top$ and $E_{\omega^0}(\cdot|\cdot)$ is the expectation taken with respect to the joint posterior density function $f(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r})$. Again, $\boldsymbol{H}_\varrho$ can be approximated by

$$\begin{aligned}\boldsymbol{H}_\varrho \approx \ddot{\varrho}(1)\Bigg[&\frac{1}{S_0}\sum_{s=1}^{S_0}\{\partial_\omega \log f(\boldsymbol{y},\boldsymbol{r},\boldsymbol{x}_{\text{mis}}^{(s)},\boldsymbol{z}^{(s)},\boldsymbol{v}^{(s)},\boldsymbol{\theta}^{(s)}|\boldsymbol{x}_{\text{obs}},\boldsymbol{\omega}^0)\}^{\otimes 2}\\ &-\left(\frac{1}{S_0}\sum_{s=1}^{S_0}\partial_\omega \log f(\boldsymbol{y},\boldsymbol{r},\boldsymbol{x}_{\text{mis}}^{(s)},\boldsymbol{z}^{(s)},\boldsymbol{v}^{(s)},\boldsymbol{\theta}^{(s)}|\boldsymbol{x}_{\text{obs}},\boldsymbol{\omega}^0)\right)^{\otimes 2}\Bigg],\end{aligned}$$

where observations $\{(\boldsymbol{x}_{\text{mis}}^{(s)},\boldsymbol{z}^{(s)},\boldsymbol{v}^{(s)},\boldsymbol{\theta}^{(s)}) : s = 1,\ldots,S_0\}$ are generated from the joint posterior distribution $f(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r})$ via the above presented Gibbs sampler.

Example 5 (Posterior Mean Distance)**.** Let $\boldsymbol{M}_d(\boldsymbol{\omega}^0) = \int d(\boldsymbol{\theta})f(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r})d\boldsymbol{x}_{\text{mis}}d\boldsymbol{z}d\boldsymbol{v}d\boldsymbol{\theta}$ and $\boldsymbol{M}_d(\boldsymbol{\omega}) = \int d(\boldsymbol{\theta})f(\boldsymbol{x}_{\text{mis}},\boldsymbol{z},\boldsymbol{v},\boldsymbol{\theta}|\boldsymbol{D}_{\text{obs}},\boldsymbol{r},\boldsymbol{\omega})d\boldsymbol{x}_{\text{mis}}d\boldsymbol{z}d\boldsymbol{v}d\boldsymbol{\theta}$ be the posterior means of $d(\boldsymbol{\theta})$ before and after introducing $\boldsymbol{\omega}$, respectively, where $d(\boldsymbol{\theta})$ is some known function of $\boldsymbol{\theta}$ of interest. Cook's distance of posterior mean of $d(\boldsymbol{\theta})$ for characterizing the effect of $\boldsymbol{\omega}$ is defined as

$$CM_d(\boldsymbol{\omega}) = \{\boldsymbol{M}_d(\boldsymbol{\omega}) - \boldsymbol{M}_d(\boldsymbol{\omega}^0)\}^\top\boldsymbol{G}_d\{\boldsymbol{M}_d(\boldsymbol{\omega}) - \boldsymbol{M}_d(\boldsymbol{\omega}^0)\},$$

where $\boldsymbol{G}_d = [\text{var}\{d(\boldsymbol{\theta})|\boldsymbol{D}_{\text{obs}},\boldsymbol{r}\}]^{-1}$. If we take $f(\boldsymbol{\omega}) = CM_d(\boldsymbol{\omega})$, it is easily shown that $\boldsymbol{\nabla}_d = 0$ and $\boldsymbol{H}_d = \boldsymbol{M}_d^{*\top}\boldsymbol{G}_d\boldsymbol{M}_d^*$, where $\boldsymbol{M}_d^* = \text{cov}_{\omega^0}\{d(\boldsymbol{\theta}),\dot{\ell}_\varrho(\boldsymbol{\omega}^0)|\boldsymbol{D}_{\text{obs}},\boldsymbol{r}\}$, and

$\mathrm{cov}_{\omega^0}(\varsigma_1, \varsigma_2)$ is the covariance of random variables ς_1 and ς_2 taken with respect to the joint posterior density $f(\boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{z}, \boldsymbol{v}, \boldsymbol{\theta}|\boldsymbol{D}_{\mathrm{obs}}, \boldsymbol{r})$. Again, $\boldsymbol{M}_d^*$ can be approximated by

$$\begin{aligned}\boldsymbol{M}_d^* \approx & \frac{1}{S_0}\sum_{s=1}^{S_0}\left\{d(\boldsymbol{\theta}^{(s)})\partial_\omega \log f(\boldsymbol{y}, \boldsymbol{r}, \boldsymbol{x}_{\mathrm{mis}}^{(s)}, \boldsymbol{z}^{(s)}, \boldsymbol{v}^{(s)}, \boldsymbol{\theta}^{(s)}|\boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{\omega}^0)\right\} \\ & - \left\{\frac{1}{S_0}\sum_{s=1}^{S_0} d(\boldsymbol{\theta}^{(s)})\right\}\left\{\frac{1}{S_0}\sum_{s=1}^{S_0}\partial_\omega \log f(\boldsymbol{y}, \boldsymbol{r}, \boldsymbol{x}_{\mathrm{mis}}^{(s)}, \boldsymbol{z}^{(s)}, \boldsymbol{v}^{(s)}, \boldsymbol{\theta}^{(s)}|\boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{\omega}^0)\right\},\end{aligned}$$

where observations $\{(\boldsymbol{x}_{\mathrm{mis}}^{(s)}, \boldsymbol{z}^{(s)}, \boldsymbol{v}^{(s)}, \boldsymbol{\theta}^{(s)}) : s = 1, \ldots, S_0\}$ are generated from the joint posterior distribution $f(\boldsymbol{x}_{\mathrm{mis}}, \boldsymbol{v}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{D}_{\mathrm{obs}}, \boldsymbol{r})$ via the above presented Gibbs sampler.

6 Simulation studies

In this section, several simulations are conducted to investigate the finite sample performance of the above proposed Bayesian methodologies.

Simulation 1. In this simulation, we consider a quantile regression model with mixed discrete and continuous covariates: $y_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \epsilon_i$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$, $\boldsymbol{x}_i = (1, x_{i1}, x_{i2})^\top$ and ϵ_i is the random error for $i = 1, \cdots, n$. Here, x_{i1} is simulated from a binomial distribution $B(1, \exp(\alpha_1)/(1+\exp(\alpha_1)))$, and x_{i2} is simulated from a normal distribution $N(\alpha_{20} + \alpha_{21}x_{i1}, \alpha_{22})$. The true values of parameters $\boldsymbol{\beta}$, α_1 and $\boldsymbol{\alpha}_2 = (\alpha_{20}, \alpha_{21}, \alpha_{22})^\top$ are taken as $\boldsymbol{\beta} = (0.5, 0.5, 0.5)^\top$, $\alpha_1 = 1$, $\boldsymbol{\alpha}_2 = (0.5, 0.5, 1)^\top$, respectively. To investigate the effect of random error distribution on parameter estimation, we consider the following four distributions for ϵ_i: (C1) ϵ_i is distributed as the standard normal distribution $N(0,1)$, (C2) ϵ_i follows a mixture of two normals $0.9N(0,1) + 0.1N(0,5)$, (C3) ϵ_i follows the t-distribution with three degrees of freedom (i.e., $\epsilon_i \sim t(3)$), and (C4) $\epsilon_i = 0.5\varepsilon_i$, where ε_i is distributed as the chi-squared distribution with three degrees of freedom (i.e., $\varepsilon_i \sim \chi^2(3)$). Thus, the τth conditional quantile of y_i is $Q_{y_i}(\tau|\boldsymbol{x}_i) = \boldsymbol{x}_i^\top\boldsymbol{\beta}_\tau$, where $\boldsymbol{\beta}_\tau = (\beta_{\tau 0}, \beta_{\tau 1}, \beta_{\tau 2})^\top$ in which $\beta_{\tau 0} = \beta_0 + Q_1(\tau)$, $\beta_{\tau 1} = \beta_1$ and $\beta_{\tau 2} = \beta_2$. For the error distributions (C1)–(C4), $Q_1(\tau)$ is the τth quantile of $N(0,1)$, the mixture of normals $0.9N(0,1) + 0.1N(0,5)$, $t(3)$, and $0.5\chi^2(3)$, respectively.

Here we assume that y_i's are completely observed, whilst x_{i1} and x_{i2} are subject to nonignorable missingness. The missing indicators r_{i1} and r_{i2} are created from the following probit regression models

$$\begin{aligned}&\Phi^{-1}\{\Pr(r_{i1} = 1)|x_{i1}, \gamma_1\} = \gamma_1 x_{i1}, \\ &\Phi^{-1}\{\Pr(r_{i2} = 1)|x_{i1}, x_{i2}, r_{i1}, \boldsymbol{\gamma}_2\} = \gamma_{21}x_{i1} + \gamma_{22}x_{i2} + \gamma_{23}r_{i1},\end{aligned}$$

respectively, where $\boldsymbol{\gamma}_2 = (\gamma_{21}, \gamma_{22}, \gamma_{23})^\top$. The true values of parameters γ_1 and $\boldsymbol{\gamma}_2$ are taken as $\gamma_1 = 0.7$ and $\boldsymbol{\gamma}_2 = (0.5, 0.1, 0.1)^\top$, respectively. The average proportions of missing data for x_{i1} and x_{i2} for $n = 500$ together with 200 replications are about 31% and 30%, respectively.

To investigate the sensitivity of Bayesian estimation to prior inputs, we consider the following three types of hyperparameters.

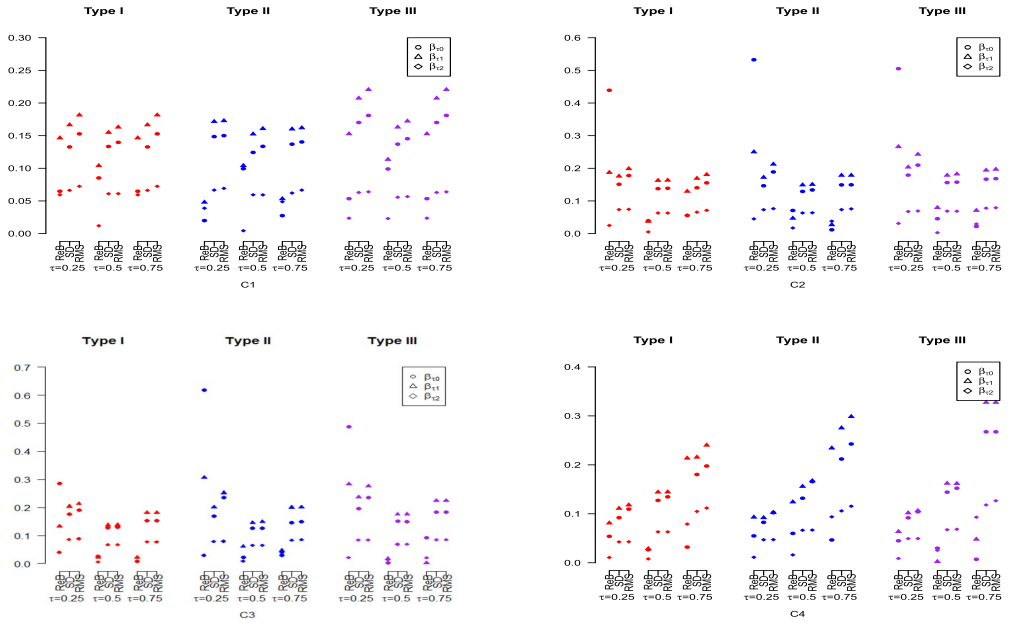

Figure 1: Performance of parameters $\boldsymbol{\beta}_\tau$ in Experiment 1.

Type I: The hyperparameters $\boldsymbol{\beta}_\tau^0$, $\boldsymbol{\alpha}_1^0$, $\boldsymbol{\alpha}_2^0$, $\boldsymbol{\gamma}_1^0$ and $\boldsymbol{\gamma}_2^0$ are taken to be their corresponding true values; $\boldsymbol{\Sigma}_{\tau\beta}^0 = 0.25\boldsymbol{I}$, $\boldsymbol{\Sigma}_{\alpha k}^0 = 0.25\boldsymbol{I}$ and $\boldsymbol{\Sigma}_{\gamma j}^0 = 0.25\boldsymbol{I}$; $\alpha_\sigma^0 = 1$ and $\beta_\sigma^0 = 1$. This can be treated as a case with good prior distribution.

Type II: The hyperparameters $\boldsymbol{\beta}_\tau^0$, $\boldsymbol{\alpha}_1^0$, $\boldsymbol{\alpha}_2^0$, $\boldsymbol{\gamma}_1^0$ and $\boldsymbol{\gamma}_2^0$ are taken to be 2 times their corresponding true values; while other hyperparameters are taken to be those given in Type I. This can be regarded as a case with misspecified prior information.

Type III: The hyperparameters for $\boldsymbol{\beta}_\tau^0$, $\boldsymbol{\alpha}_1^0$, $\boldsymbol{\alpha}_2^0$, $\boldsymbol{\gamma}_1^0$ and $\boldsymbol{\gamma}_2^0$ are taken to be 10 times their corresponding true values; while $\boldsymbol{\Sigma}_{\tau\beta}^0 = 10\boldsymbol{I}$, $\boldsymbol{\Sigma}_{\alpha k}^0 = 10\boldsymbol{I}$ and $\boldsymbol{\Sigma}_{\gamma j}^0 = 10\boldsymbol{I}$. This can be treated as a case with noninformative prior information.

For each of 200 datasets generated above together with each of three hyperparameters, the preceding introduced Gibbs sampler is adopted to evaluate Bayesian estimates of unknown parameters $\boldsymbol{\beta}_\tau$, $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_1$. To investigate convergence of Gibbs sampler algorithm, we calculate the EPSR values of parameters based on three parallel series of observations simulated from three different starting values of parameters. For several runs tested, we observe that the EPSR values of parameters are less than 1.2 after about 1500 iterations. Thus, $\mathcal{T} = 7000$ observations are collected after 3000 burn-in iterations to evaluate Bayesian estimates for each of 200 replications. To save space, only results of $\boldsymbol{\beta}_\tau$ for three quantiles (i.e., $\tau = 0.25$, $\tau = 0.5$ and $\tau = 0.75$) are reported in Figure 1. Results for $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_j$ are given in Table S1 in Supplementary Materials. Examination of Figure 1 and Table S1 shows that (i) Bayesian estimates evaluated with Type I prior are better than those obtained with Type II and Type III

priors but their differences are quite small; (ii) Bayesian estimates obtained with Type I and Type II priors are better than those obtained with Type III prior; (iii) "ReB"(i.e., "Relative Bias") values are less than 0.4, and "SD" and "RMS" values are less than 0.3, and "SD" values are quite close to those of "RMS" regardless of any priors and error distributions and quantiles. These findings evidence that the proposed Bayesian estimates are quite accurate, and not sensitive to prior inputs and error distributions and quantiles. In particular, these empirical results indicate that the proposed Bayesian estimation approach is valid in the case of a regression model with light tail distribution errors. Also, we calculate the corresponding results for $n = 200$. To save space, we omit them. Comparing the results for $n = 200$ and 500, we observe that "SD" and "RMS" values decrease as the sample size increases.

Simulation 2. In this simulation, the preceding proposed Bayesian variable selection procedure is employed to simultaneously select explanatory variables in response model, covariate model and missingness data mechanism model. To this end, response variables y_i's are generated by $y_i = \beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta} + \eta x_{3i}\epsilon_i$, where $\beta_0 = 1$ is a fixed value, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{10})^\top$, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{i,10})^\top$ for $i = 1, \ldots, n$ with $n = 500$, η is some fixed value (e.g., $\eta = 0.5$) and ϵ_i's follow the standard normal distribution. Here x_{i1} and x_{i2} are simulated from the following normal distributions: $N(\alpha_{10} + \alpha_{11}x_{i3} + \ldots + \alpha_{18}x_{i,10}, \alpha_{19})$ and $N(\alpha_{20} + \alpha_{21}x_{i1} + \alpha_{22}x_{i3} + \ldots + \alpha_{29}x_{i,10}, \alpha_{2,10})$, respectively, where $\alpha_{10} = -0.5$ and $\alpha_{20} = 1.0$ are regarded as fixed parameters for identification; x_{i3} is simulated from $U(1, 5)$; and $x_{i4}, \ldots, x_{i,10}$ are independently simulated from $N(0, 1)$. The true values of parameters α_{19}, $\alpha_{2,10}$, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_1 = (\alpha_{11}, \alpha_{11}, \ldots, \alpha_{18})^\top$ and $\boldsymbol{\alpha}_2 = (\alpha_{21}, \alpha_{21}, \ldots, \alpha_{29})^\top$ are taken to be $\alpha_{19} = 1.0$, $\alpha_{2,10} = 1.0$, $\boldsymbol{\beta} = (-1, 1, -1, 1, 0, \ldots, 0)^\top$, $\boldsymbol{\alpha}_1 = (0.5, 0, \ldots, 0)^\top$, and $\boldsymbol{\alpha}_2 = (0.5, -0.5, 0, \ldots, 0)^\top$, respectively, which show that there are 4 nonzero coefficients in $\boldsymbol{\beta}$, 1 nonzero coefficient in $\boldsymbol{\alpha}_1$, and 2 nonzero coefficients in $\boldsymbol{\alpha}_2$. Similarly, the τth conditional quantile of y_i is $Q_{y_i}(\tau|\boldsymbol{x}_i) = \beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau$, where $\boldsymbol{\beta}_\tau = (\beta_{\tau 1}, \ldots, \beta_{\tau,10})^\top$ with $\beta_{\tau 3} = \beta_3 + Q_1(\tau)$ in which $Q_1(\tau)$ is η times the τth quantile of $N(0, 1)$, and $\beta_{\tau j} = \beta_j$ for $j = 1, 2, 4, \ldots, 10$. It is assumed that y_i and $x_{i3}, \ldots, x_{i,10}$ are completely observed, while x_{i1} and x_{i2} are subject to missingness. The missing indicators for x_{i1} and x_{i2} are created by (2.12) with $\mu_{i1} = \gamma_{10} + \gamma_{11}x_{i1} + \gamma_{12}x_{i3} + \ldots + \gamma_{19}x_{i,10}$ and $\mu_{i2} = \gamma_{20} + \gamma_{21}x_{i1} + \ldots + \gamma_{2,10}x_{i,10}$, respectively, where $\gamma_{10} = -1.5$ and $\gamma_{20} = 1.5$ are treated as fixed parameters for identification. The true values of parameters $\boldsymbol{\gamma}_1 = (\gamma_{11}, \ldots, \gamma_{1,9})^\top$ and $\boldsymbol{\gamma}_2 = (\gamma_{21}, \ldots, \gamma_{2,10})^\top$ are taken to be $\boldsymbol{\gamma}_1 = (0.5, 0.5, 0, \ldots, 0)^\top$ and $\boldsymbol{\gamma}_2 = (0.5, -0.5, -0.5, 0, \ldots, 0)^\top$, respectively, which indicate that there are 2 nonzero coefficients in $\boldsymbol{\gamma}_1$ and 3 nonzero coefficients in $\boldsymbol{\gamma}_2$. The preceding introduced Bayesian variable selection procedure together with the hyperparameters $\delta_\beta = 0.1$, $\delta_\alpha = 0.1$ and $\delta_\gamma = 0.1$ are used to identify nonzero components in $\boldsymbol{\beta}_\tau$, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_j$ for $k, j = 1, 2$ based on 10000 observations collected after 10000 burn-in iterations for each of 100 replications. Also, to investigate the performance of the proposed Bayesian variable selection procedure, we calculate the L_2 norm between the estimate and its true value (e.g., $L_2 = (\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^\top(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)$), and the mean square error (e.g., MSE $= (\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^\top \boldsymbol{X}^\top \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)/n$) of parameter vector of interest (e.g., $\boldsymbol{\beta}_\tau$) for each of 100 replications, where $\widehat{\boldsymbol{\beta}}_\tau$ is Bayesian estimate of $\boldsymbol{\beta}_\tau$. At the same time, we evaluate the average numbers of zero components correctly detected as zero (denoted as "C"), and nonzero components incorrectly identified as zero (denoted as "IC"). For comparison, we compute the corresponding results for Bayesian LASSO. Results for three quantiles (i.e.,

		$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
Meth.	Para.	L_2	MSE	C	IC	L_2	MSE	C	IC	L_2	MSE	C	IC
BaL	$\boldsymbol{\beta}_\tau$	0.002	0.067	5.95	0.00	0.004	0.066	5.97	0.00	0.005	0.067	5.96	0.00
	$\boldsymbol{\alpha}_1$	0.000	0.002	7.00	0.00	0.000	0.002	7.00	0.00	0.001	0.003	7.00	0.00
	$\boldsymbol{\alpha}_2$	0.000	0.009	7.00	0.00	0.001	0.010	6.98	0.00	0.010	0.012	6.99	0.01
	$\boldsymbol{\gamma}_1$	0.005	0.021	6.97	0.22	0.000	0.022	6.99	0.09	0.006	0.021	6.99	0.03
	$\boldsymbol{\gamma}_2$	0.009	0.026	6.99	0.30	0.000	0.030	7.00	0.23	0.004	0.030	7.00	0.07
BL	$\boldsymbol{\beta}_\tau$	0.003	0.106	5.57	0.00	0.015	0.085	5.77	0.00	0.009	0.119	5.52	0.00
	$\boldsymbol{\alpha}_1$	0.000	0.016	6.78	0.00	0.000	0.015	6.86	0.00	0.001	0.019	6.69	0.00
	$\boldsymbol{\alpha}_2$	0.000	0.025	6.74	0.00	0.004	0.027	6.81	0.00	0.014	0.028	6.78	0.00
	$\boldsymbol{\gamma}_1$	0.004	0.040	6.78	0.01	0.002	0.049	6.79	0.00	0.008	0.048	6.82	0.00
	$\boldsymbol{\gamma}_2$	0.015	0.055	6.80	0.14	0.005	0.052	6.77	0.04	0.005	0.055	6.86	0.00

Note: 'BaL' represents the Bayesian adaptive LASSO, that is our proposed method, 'BL' represents the Bayesian LASSO, the compared method, 'C' represents the average number of zero components correctly identified as zero, and 'IC' is the average number of nonzero components incorrectly identified as zero.

Table 1: Performance of Bayesian variable selection in Experiment 2.

$\tau = 0.25$, 0.5 and 0.75) are reported in Table 1. Inspection of Table 1 shows that (i) the values of IC's are quite close to zero, and the values of C's are rather close to the true numbers of zero elements for all the parameters as expected; (ii) the values of L_2 for all the parameters are less than 0.02, which is quite close to zero, regardless of the considered three quantiles; (iii) the values of MSE's for all the parameters are less than 0.12 regardless of the considered three quantiles; (iv) the proposed variable selection procedure has smaller L_2 norms and MSE values than Bayesian LASSO. These findings indicate that the proposed Bayesian variable selection procedure performs well and has better performance than Bayesian LASSO.

Simulation 3. To illustrate the above introduced Bayesian local influence measures in detecting influential observations and incorrectly specified missingness data mechanism, we consider a linear regression model: $y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i$, where $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, x_{i3})^\top$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$, and ϵ_i follows the standard normal distribution $N(0,1)$ for $i = 1, \ldots, n$. Here, x_{i1}, x_{i2} and x_{i3} are independently generated from the following normal distributions $N(\alpha_{10}, \alpha_{11})$, $N(\alpha_{20} + \alpha_{21} x_{i1}, \alpha_{22})$ and $N(\alpha_{30} + \alpha_{31} x_{i1} + \alpha_{32} x_{i2}, \alpha_{33})$, respectively. The true values of parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_1 = (\alpha_{10}, \alpha_{11})^\top$, $\boldsymbol{\alpha}_2 = (\alpha_{20}, \alpha_{21}, \alpha_{22})^\top$ and $\boldsymbol{\alpha}_3 = (\alpha_{30}, \alpha_{31}, \alpha_{32}, \alpha_{33})^\top$ are taken as $\boldsymbol{\beta} = (0.5, 0.5, 0.5, 0.5)^\top$, $\boldsymbol{\alpha}_1 = (0.5, 1.0)^\top$, $\boldsymbol{\alpha}_2 = (0.5, 0.05, 1.0)^\top$ and $\boldsymbol{\alpha}_3 = (0.5, 0.05, 0.05, 1.0)^\top$, respectively. Again, the τth conditional quantile of y_i is $Q_{y_i}(\tau | \boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau$, where $\boldsymbol{\beta}_\tau = (\beta_{\tau 0}, \beta_{\tau 1}, \beta_{\tau 2}, \beta_{\tau 3})^\top$ with $\beta_{\tau 0} = \beta_0 + Q_1(\tau)$ in which $Q_1(\tau)$ is the τth quantile of $N(0,1)$, $\beta_{\tau j} = \beta_j$ for $j = 1, 2, 3$. Here we assume that y_i's are completely observed, while x_{i1}, x_{i2} and x_{i3} are subject to nonignorable missingness. The missing indicators r_{i1}, r_{i2} and r_{i3} for covariates x_{i1}, x_{i2} and x_{i3} are created from the following probit regression models

$$\begin{aligned}
&\Phi^{-1}\{\Pr(r_{i1} = 1 | x_{i1}, \boldsymbol{\gamma}_1)\} = \gamma_{10} + \gamma_{11} x_{i1}, \\
&\Phi^{-1}\{\Pr(r_{i2} = 1 | x_{i1}, x_{i2}, \boldsymbol{\gamma}_2)\} = \gamma_{20} + \gamma_{21} x_{i1} + \gamma_{22} x_{i2}, \\
&\Phi^{-1}\{\Pr(r_{i3} = 1 | x_{i1}, x_{i2}, x_{i3}, \boldsymbol{\gamma}_3)\} = \gamma_{30} + \gamma_{31} x_{i1} + \gamma_{32} x_{i2},
\end{aligned}$$

respectively, where $\boldsymbol{\gamma}_1 = (\gamma_{10}, \gamma_{11})^\top$, $\boldsymbol{\gamma}_2 = (\gamma_{20}, \gamma_{21}, \gamma_{22})^\top$, and $\boldsymbol{\gamma}_3 = (\gamma_{30}, \gamma_{31}, \gamma_{32})^\top$. The true values of parameters $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ and $\boldsymbol{\gamma}_3$ are taken as $\boldsymbol{\gamma}_1 = (0.5, 0.5)^\top$, $\boldsymbol{\gamma}_2 = (0.5, 0.1, 0.3)^\top$ and $\boldsymbol{\gamma}_3 = (0.5, 0.1, 0.1)^\top$, respectively. Missingness data mechanisms for x_{i1} and x_{i2} are nonignorable, while x_{i3} is subject to MAR. Five influential cases are created by changing y_i as $y_i + 6$ for $i = 50$, 139, 301, 313 and 437, respectively.

We consider simultaneous perturbation to the above generated data (e.g., case-weight perturbation), priors (e.g., see Example 3), missingness data mechanism (e.g., see Example 5). The log-likelihood function of the perturbed model is given by

$$\begin{aligned}\ell(\boldsymbol{\omega}) \quad = \quad & \textstyle\sum_{i=1}^{n}[-0.5\omega_i \ln(2\pi\kappa_2 v_i/\sigma) + 0.5\ln\omega_i - \omega_i\sigma(y_i - \boldsymbol{x}_i^\top\beta - \kappa_1 v_i)^2/(2\kappa_2 v_i) \\ & + r_{i3}\ln\{\Phi(\chi_{\gamma i3}^{\omega_x})\} + (1 - r_{i3})\ln\{1 - \Phi(\chi_{\gamma i3}^{\omega_x})\}] + \ell(\boldsymbol{\omega}_\beta, \omega_\sigma, \boldsymbol{\omega}_\gamma)\end{aligned}$$

in which $\ell(\boldsymbol{\omega}_\beta, \omega_\sigma, \boldsymbol{\omega}_\gamma)$ has the form of

$$\begin{aligned}\ell(\boldsymbol{\omega}_\beta, \omega_\sigma, \boldsymbol{\omega}_\gamma) \quad = \quad & \frac{p}{2}\ln\omega_\beta + \ln|\boldsymbol{\Sigma}_{\tau\beta}^0| - \frac{\omega_\beta}{2}(\boldsymbol{\beta}_\tau - \boldsymbol{\beta}_\tau^0)^\top {\boldsymbol{\Sigma}_{\tau\beta}^0}^{-1}(\boldsymbol{\beta}_\tau - \boldsymbol{\beta}_\tau^0) \\ & + \frac{p_{\gamma_1}}{2}\ln\omega_{\gamma_1} + \ln|\boldsymbol{\Sigma}_{\gamma_1}^0| - \frac{\omega_{\gamma_1}}{2}(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_1^0)^\top {\boldsymbol{\Sigma}_{\gamma_1}^0}^{-1}(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_1^0) \\ & + \frac{p_{\gamma_2}}{2}\ln\omega_{\gamma_2} + \ln|\boldsymbol{\Sigma}_{\gamma_2}^0| - \frac{\omega_{\gamma_2}}{2}(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2^0)^\top {\boldsymbol{\Sigma}_{\gamma_2}^0}^{-1}(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2^0) \\ & + \frac{p_{\gamma_3}}{2}\ln\omega_{\gamma_3} + \log|\boldsymbol{\Sigma}_{\gamma_3}^0| - \frac{\omega_{\gamma_3}}{2}(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_3^0)^\top {\boldsymbol{\Sigma}_{\gamma_3}^0}^{-1}(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_3^0) \\ & + \alpha_\sigma^0\ln(\omega_\sigma\beta_\sigma^0) - \ln\Gamma(\alpha_\sigma^0) + (\alpha_\sigma^0 - 1)\ln\sigma - \omega_\sigma\beta_\sigma^0\sigma,\end{aligned} \tag{6.1}$$

where $\chi_{\gamma i3}^{\omega_x} = \gamma_{30} + \gamma_{31}x_{i1} + \gamma_{32}x_{i2} + \omega_x x_{i3}$, and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n, \omega_x, \omega_\beta, \omega_\sigma, \omega_{\gamma_1}, \omega_{\gamma_2}, \omega_{\gamma_3})^\top$. In this case, $\boldsymbol{\omega}^0 = (1.0, \cdots, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 1.0)^\top$ represents no perturbation. Again, the aforementioned Gibbs sampler together with Type III prior for parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ is adopted to estimate parameters, and calculate $\boldsymbol{G}(\boldsymbol{\omega}^0)$ and Bayesian local influence measures corresponding to Bayes factor (i.e., FIC_{B,e_j}), Kullback-Leibler divergence (i.e., $\mathrm{SIC}_{D_\varrho,e_j}$) and posterior mean distance (i.e., SIC_{M_h,e_j}) of $d(\boldsymbol{\theta}) = \boldsymbol{\theta}$ based on $S_0 = 7000$ observations collected after 3000 burn-in iterations. Results for $\tau = 0.5$ are presented in Figure 2. Examination of Figure 2 indicates that cases 50, 139, 301, 313 and 437 are detected to be influential, and the incorrectly specified missingness data mechanism is identified to have a large effect as expected.

7 An example

To illustrate the above proposed methodologies, we consider the US News College data, which are from the 1995 US News report on American colleges and universities. The dataset is available at the website http://lib.stat.cmu.edu/datasets/colleges. The main interest is to investigate the relationship between the population quantile of the ratio of graduating seniors to number enrolling four years earlier (GRADRAT, y) and some factors, for example, Average Combined SAT score (ACS, x_1), Average ACT score (AAS, x_2), natural logarithm of number of applications received (LNR, x_3), natural logarithm of number of applicants accepted (LNA, x_4), natural logarithm of num-

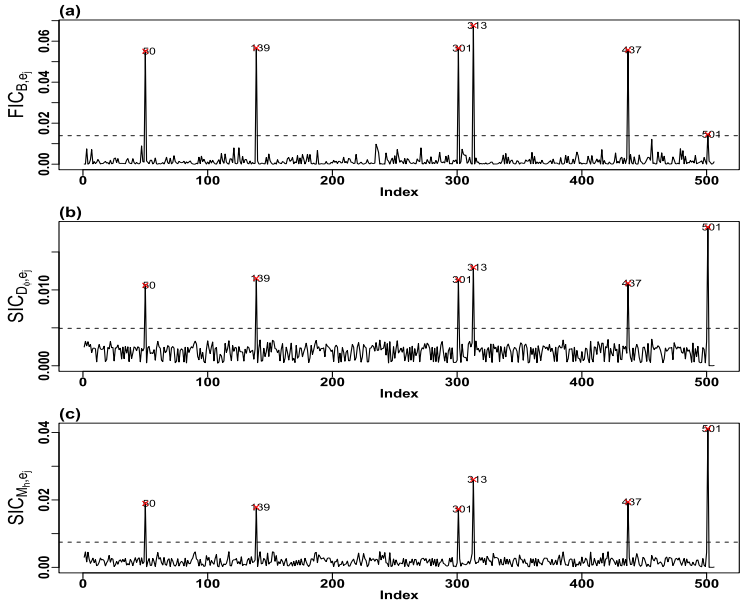

Figure 2: Index plots of (a) FIC_{B,e_j}, (b) SIC_{D_ϱ,e_j} and (c) SIC_{M_h,e_j} under incorrect missing mechanism in Experiment 3.

ber of new students enrolled (LNE, x_5), room and board costs (RBC, x_6), natural logarithm of instructional expenditure per student (LIE, x_7), ratio of student to faculty (RSF, x_8). The dataset includes 1203 observations except for a false case whose GRADRAT bigger than 100%, and variables y and x_8 are completely observed, while covariates $x_1, \ldots, x_7$ are subject to missingness and their corresponding missing proportions are 39.2%, 45.2%, 0.50%, 0.42%, 0.17%, 4.90%, and 2.08%, respectively. To roughly unify the scales, the raw data are standardized on the basis of the fully observed data.

As an illustration of the preceding introduced methodologies, we consider the following conditional QR model: $Q_\tau(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}_\tau) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau$ for $i = 1, \ldots, 1203$, where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{i8})^\top$, and $\boldsymbol{\beta}_\tau = (\beta_{\tau 1}, \ldots, \beta_{\tau 8})^\top$. It is assumed that $x_{ik}|\boldsymbol{\alpha}_k \sim N(\alpha_{k1}x_{i1} + \ldots + \alpha_{k,k-1}x_{i,k-1} + \alpha_{kk}x_{i8}, \alpha_{k,k+1})$, where $\boldsymbol{\alpha}_k = (\alpha_{k1}, \ldots, \alpha_{k,k+1})^\top$ for $k = 1, \ldots, 7$, and its corresponding missingness data mechanism is specified by the Probit model defined in (2.12) with $\boldsymbol{x}_{zij} = (x_{i1}, \ldots, x_{ij}, x_{i8})^\top$ and $\boldsymbol{\gamma}_j = (\gamma_{j1}, \ldots, \gamma_{j,j+1})^\top$ for $j = 1, \ldots, 7$.

The considered dataset is equally divided into two parts, i.e., the training dataset used to select important variables, and the testing dataset adopted to estimate parameters and make Bayesian local influence analysis for the selected model. The aforementioned Bayesian variable selection procedure together with noninformative prior specification of parameters (e.g., $\sigma \sim \Gamma(1, 1)$, $\delta_\beta = 0.1$, $\delta_\alpha = 0.1$ and $\delta_\gamma = 0.1$) is used to estimate parameters in $\boldsymbol{\beta}_\tau$, select significant covariates in QR model, explanatory variables associated with the distribution of missing covariates and missingness

data mechanism model. To monitor convergence of the Gibbs sampler, we calculate the EPSR values of all unknown parameters based on three different starting values of parameters. To save space, the EPSR values of parameters against iterations for training dataset are shown in Figure S1 in Supplementary Materials. Inspection of Figure S1 shows that the algorithm attains convergence after about 13,000 iterations. To this end, we collect 15,000 observations after 15,000 burn-in iterations to evaluate Bayesian estimates and select significant explanatory variables for $\tau = 0.5$. Bayesian estimates and 95% confidence intervals (CI) of unknown parameters for training dataset are presented in Table S4. Inspection of Tables S4 indicates that (i) x_3, x_4, x_5 and x_7 have little effect on y yielding $Q_{0.5}(y|x) = 0.151x_1 + 0.336x_2 + 0.144x_6 - 0.077x_8$; (ii) missingness data mechanisms for x_1 and x_2 are nonignorable, while missingness data mechanisms for x_3, x_4, x_5, x_6 and x_7 are missing at random.

To illustrate Bayesian local influence measures introduced above, we consider simultaneous perturbation to the testing data, priors and missingness data mechanisms, whose log-likelihood function is given by

$$\begin{aligned}\ell(\boldsymbol{\omega}) \;=\; & \sum_{i=1}^{n}\Big[-\frac{\omega_i}{2}\ln(2\pi\kappa_2 v_i/\sigma)+\frac{1}{2}\ln\omega_i-\frac{\omega_i\sigma(y_i-\beta_{\tau 0}-\boldsymbol{x}_i^{\top}\boldsymbol{\beta}_\tau-\kappa_1 v_i)^2}{2\kappa_2 v_i}\\ & +r_{i1}\ln\{\Phi(\chi_{\gamma i1}^{\omega_x})\}+(1-r_{i1})\ln\{1-\Phi(\chi_{\gamma i1}^{\omega_x})\}+r_{i2}\ln\{\Phi(\chi_{\gamma i2}^{\omega_x})\}\\ & +(1-r_{i2})\ln\{1-\Phi(\chi_{\gamma i2}^{\omega_x})\}\Big]+\ell(\boldsymbol{\omega}_\beta,\omega_\sigma,\boldsymbol{\omega}_\gamma),\end{aligned}$$

where $\ell(\boldsymbol{\omega}_\beta,\omega_\sigma,\boldsymbol{\omega}_\gamma)$ is defined in (6.1), and $\chi_{\gamma i1}^{\omega_x}=\omega_{x1}\gamma_{11}x_{i1}$, $\chi_{\gamma i2}^{\omega_x}=\omega_{x2}\gamma_{22}x_{i2}$, $\boldsymbol{\Sigma}_{\tau\beta}^0=\mathrm{diag}(\nu_1^\beta,\ldots,\nu_p^\beta)$, $\boldsymbol{\Sigma}_{\gamma j}^0=\mathrm{diag}(\nu_{j1}^\gamma,\ldots,\nu_{j,j+1}^\gamma)$ in which ν_j^β and ν_{jl}^γ for $l=1,2,3$ are parameters for variable selection. Here missingness data mechanisms for x_1 and x_2 are defined in (2.12) with $\boldsymbol{x}_{zij}=x_{ij}$ and $\boldsymbol{\gamma}_j=\gamma_{jj}$ for $j=1,2$, and $\boldsymbol{x}_{zij}=(x_{i1},x_{i2})^{\top}$ and $\boldsymbol{\gamma}_j=(\gamma_{j1},\gamma_{j2})^{\top}$ for $j=3,\ldots,7$. We take the hyperparameters $\boldsymbol{\beta}_\tau^0$, $\boldsymbol{\alpha}_k^0$ and $\boldsymbol{\gamma}_j^0$ as their corresponding Bayesian estimates given in Table S4, and set $\alpha_\sigma^0=1.0$, $\beta_\sigma^0=1.0$, $\boldsymbol{\Sigma}_{\alpha k}^0=\mathrm{diag}(\nu_{k1}^\alpha,\ldots,\nu_{k,k}^\alpha)$ for $k=1,\cdots,7$, and $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_n,\omega_{x1},\omega_{x2},\omega_\beta,\omega_{\gamma_1},\omega_{\gamma_2},\omega_{\gamma_3},\omega_\sigma)^{\top}$. In this case, $\omega^0=(1.0,\ldots,1.0)^{\top}$ represents no perturbation.

Again, the aforementioned Gibbs sampler algorithm is used to calculate $\boldsymbol{G}(\boldsymbol{\omega}^0)$ and Bayesian local influence measures corresponding to Bayes factor (i.e., FIC_{B,e_j}), Kullback-Leibler divergence (i.e., $\mathrm{SIC}_{D_\varrho,e_j}$) and posterior mean distance (i.e., SIC_{M_h,e_j}) of $d(\boldsymbol{\theta})=\boldsymbol{\theta}$ for the testing dataset based on $S_0=15,000$ observations collected after 15,000 burn-in iterations. Results are given in Figure 3. Examination of Figure 3 indicates that cases 212, 590 and 595 are identified as influential by FIC_{B,e_i}, $\mathrm{SIC}_{D_\phi,e_i}$ and SIC_{M_d,e_i}, while case 544 is detected as influential by FIC_{B,e_i} and $\mathrm{SIC}_{D_\phi,e_i}$, and the priors of $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\gamma}_k$ for $k=1,2,3$ are not detected as inappropriate. In particular, missingness data mechanisms for x_1 and x_2 (i.e., cases 603 and 604) are detected to have little effect, which indicates that missingness data mechanism for x_1 and x_2 are missing not at random (MNAR). These results are consistent with those with Bayesian variable selection approach, which show that Bayesian local influence analysis method can be used to detect the misspecification of the posited missingness data mechanism model.

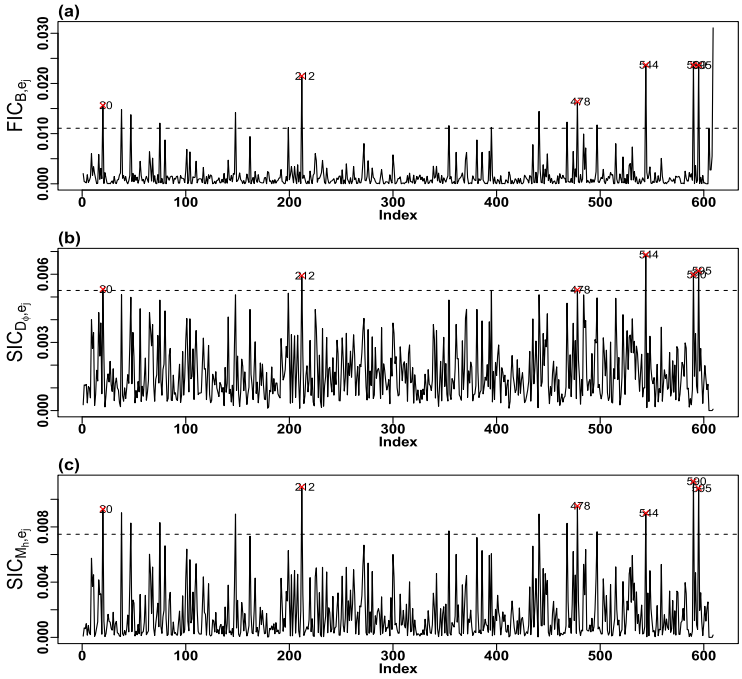

Figure 3: Index plots of (a) FIC_{B,e_j}, (b) SIC_{D_ϱ,e_j} and (c) SIC_{M_h,e_j} in real example.

	With		Without	
Par.	Est	CI	Est	CI
$\beta_{\tau 1}$	0.289	(0.206,0.372)	0.297	(0.214,0.379)
$\beta_{\tau 2}$	0.233	(0.169,0.298)	0.233	(0.170,0.295)
$\beta_{\tau 6}$	0.074	(-0.001,0.164)	0.072	(-0.001,0.161)
$\beta_{\tau 8}$	-0.125	(-0.202,-0.048)	-0.118	(-0.194,-0.040)
γ_{11}	0.278	(0.210,0.347)	0.283	(0.217,0.349)
γ_{22}	-0.469	(-0.535,-0.407)	-0.469	(-0.534,-0.405)

Table 2: Bayesian estimates (Est) and 95% confidence intervals (CI) of parameters $\boldsymbol{\beta}_\tau$ with $\tau = 0.5$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ with and without cases 212, 544, 590 and 595 for testing dataset in real example.

To investigate the effect of individuals 212, 544, 590 and 595, we recalculated Bayesian estimates of parameters with these individuals deleted from the testing dataset. Bayesian estimates and 95% confidence intervals (CI) of parameters $\boldsymbol{\beta}_\tau$, $\boldsymbol{\gamma}_1 = \gamma_{11}$ and $\boldsymbol{\gamma}_2 = \gamma_{22}$ with and without cases 212, 544, 590 and 595 are presented in Table 2. While results for parameters $\boldsymbol{\alpha}_k$ $(k = 1, \ldots, 7)$ and $\boldsymbol{\gamma}_j$ $(j = 3, \ldots, 7)$ are given in Table S5. Inspection of Tables 2 and S5 indicates that individuals 212, 544, 590 and 595 have a relatively large effect on Bayesian estimates of parameters.

8 Conclusion

This paper considers Bayesian inference on a quantile regression model in the presence of nonignorable missing covariates. To accommodate a general case of covariates including discrete and continuous variables, we use a sequence of one-dimensional exponential family distributions to specify the joint distribution of missing covariates. In a Bayesian framework, to develop an effective and feasible sampling algorithm, we reformulate the considered quantile regression model as a hierarchical model by regarding an asymmetric Laplace distribution of response variable as a mixture of exponential and normal distributions. A probit regression model is adopted to specify missingness data mechanism, and a truncated normal latent variable is employed to generate missing indicator for covariate. Similar to the logistic regression model, maximum likelihood estimators of parameters in the considered probit model are not robust to outliers or influential observations (Pregibon, 1982), a robust robit regression model (Liu, 2004) may be adopted to specify missingness data mechanism by replacing the normal distribution in the probit regression model with a t-distribution with known or unknown degrees of freedom. A Gibbs sampler is developed to sample observations, required in evaluating Bayesian results, from the posterior distributions of parameters, missing covariates and latent variables. Empirical results evidence that the proposed Bayesian estimates are reasonably accurate, and are not sensitive to prior inputs, error distribution assumptions and quantiles with probabilities in the considered region. For the quantiles with probabilities over 0.9 or close to zero, the proposed Gibbs sampler suffers from the slow convergence and poor mixing issues, and the precision of parameter estimation gets poorer in that no data are available. In this case, one possible solution is to select an appropriate prior for σ. The Dirichlet process prior may be a good option.

A Bayesian variable selection procedure is proposed by imposing a conditional Laplace prior on unknown parameters associated with covariates in the considered quantile regression model and missingness data mechanism model. Two Bayesian local influence measures (i.e., the first- and second-order influence measures) are developed to assess the effect of minor perturbation to the data, the priors and missingness data mechanism model for any objective functions such as ϕ-divergence, posterior mean distance and Bayes factor. Empirical results evidence that the proposed Bayesian variable selection procedure and Bayesian local influence measures perform well.

Although we only consider a parametric probit regression model for missingness data mechanism, the above presented approach can be extended to a nonparametric model such as generalized additive model. For example, for $p_{ij} = \Pr(r_{ij} = 1|\boldsymbol{r}_{i(j)}, \boldsymbol{x}_{i(j)})$, we consider the following generalized additive model: $h(p_{ij}) = \varphi_0 + \varphi_1(x_{i1}) + \ldots + \varphi_j(x_{ij}) + \varphi_{j+1}(r_{i1}) + \ldots + \varphi_{2j-1}(r_{i,j-1})$, where $h(\cdot)$ is some known link function, $\varphi_l(\cdot)$'s are unknown smooth functions for $l = 1, \ldots, 2j-1$. The backfitting algorithm (Hastie and Tibshirani, 1990) or boosting approach (Schmid and Hothorn, 2008) or rank reduced approach may be employed to estimate $\varphi_l(\cdot)$ for $l = 1, \ldots, 2j-1$. In this paper, we only consider the case that covariates are subject to missingness, but response is fully observed. In fact, the preceding proposed methods are still available for the case that response and/or covariates are subject to missingness.

We investigate Bayesian local influence analysis via the approach proposed by Zhu et al. (2011). But, as the Associated Editor pointed out, the approach of Zhu et al. (2011) is strong but really quite complicated for routine usage. To address this issue, it is interesting to extend the relative entropy of Clarke and Gustafson (1998) and Fréchet derivative of the posterior density with respect to the prior distribution (Gustafson and Wasserman, 1995) to our considered QR model. We are working on this topic.

As the Associated Editor pointed out, it is interesting to consider the posterior variance distance of $d(\boldsymbol{\theta})$, i.e., $f(\boldsymbol{\omega}) = \{\boldsymbol{V}(\boldsymbol{\omega}) - \boldsymbol{V}(\boldsymbol{\omega}^0)\}^\top \boldsymbol{G}_v \{\boldsymbol{V}(\boldsymbol{\omega}) - \boldsymbol{V}(\boldsymbol{\omega}^0)\}$, where $\boldsymbol{V}(\boldsymbol{\omega})$ is the posterior variance of $d(\boldsymbol{\theta})$ (Gustafson and Clarke, 2004), and $\boldsymbol{G}_v$ is some positive definite matrix.

Supplementary Material

Supplementary Material of "Bayesian Quantile regression with mixed discrete and non-ignorable missing covariates" (DOI: 10.1214/19-BA1165SUPP; .pdf).

References

Albert, J. H. and Chib, S. (1993). "Bayesian analysis of binary and polychotomous response data." *Journal of the American Statistical Association*, 88(422): 669–679. MR1224394. 583

Alhamzawi, R., Yu, K., and Benoit, D. F. (2012). "Bayesian adaptive Lasso quantile regression." *Statistical Modelling*, 12(3): 279–297. MR3179503. doi: https://doi.org/10.1177/1471082X1101200304. 580, 586

Andrews, D. F. and Mallows, C. L. (1974). "Scale mixtures of normal distributions." *Journal of the Royal Statistical Society B*, 99–102. MR0359122. 586

Brooks, S. and Andrew, G. (1998). "General methods for monitoring convergence of iterative simulations." *Journal of Computational and Graphical Statistics*, 7(4): 434–455. MR1665662. doi: https://doi.org/10.2307/1390675. 585

Cade, B. S. and Noon, B. R. (2003). "A gentle introduction to quantile regression for ecologists." *Frontiers in Ecology and the Environment*, 1(8): 412–420. 579

Canay, I. A. (2011). "A simple approach to quantile regression for panel data." *The Econometrics Journal*, 14(3): 368–386. MR2853232. doi: https://doi.org/10.1111/j.1368-423X.2011.00349.x. 579

Casella, G. (2001). "Empirical Bayes Gibbs sampling." *Biostatistics*, 2(4): 485–500. 586

Chen, X., Wan, A. T., and Zhou, Y. (2015). "Efficient quantile regression analysis with missing observations." *Journal of the American Statistical Association*, 110(510): 723–741. MR3367260. doi: https://doi.org/10.1080/01621459.2014.928219. 580

Chernozhukov, V. (2005). "Extremal quantile regression." *Annals of Statistics*, 33(2): 806–839. MR2163160. doi: https://doi.org/10.1214/009053604000001165. 579

Clarke, B. and Gustafson, P. (1998). "On the overall sensitivity of the posterior distribution to its inputs." *Journal of Statistical Planning and Inference*, 71(1): 137–150. MR1651796. doi: https://doi.org/10.1016/S0378-3758(98)00014-7. 600

Cook, R. D. (1986). "Assessment of local influence." *Journal of the Royal Statistical Society B*, 48(1): 133–169. MR0867994. 580

Gaglianone, W. P., Lima, L. R., Linton, O., and Smith, D. R. (2011). "Evaluating value-at-risk models via quantile regression." *Journal of Business and Economic Statistics*, 29(1): 150–160. MR2789438. doi: https://doi.org/10.1198/jbes.2010.07318. 579

Gelman, A. (1996). *Inference and monitoring convergence in Markov chain Monte Carlo in practice*. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds. MR1397966. doi: https://doi.org/10.1007/978-1-4899-4485-6. 585

Geraci, M. and Bottai, M. (2006). "Quantile regression for longitudinal data using the asymmetric Laplace distribution." *Biostatistics*, 8(1): 140–154. 579

Gustafson, P. and Clarke, B. (2004). "A decomposition for the posterior variance." *Journal of Statistical Planning and Inference*, 119: 311–327. MR2019643. doi: https://doi.org/10.1016/S0378-3758(02)00491-3. 600

Gustafson, P. and Wasserman, L. (1995). "Local sensitivity diagnostics for Bayesian inference." *Annals of Statistics*, 23(6): 2153–2167. MR1389870. doi: https://doi.org/10.1214/aos/1034713652. 600

Hallock, K. F. and Koenker, R. W. (2001). "Quantile regression." *Journal of Economic Perspectives*, 15(4): 143–156. MR2268657. doi: https://doi.org/10.1017/CBO9780511754098. 579

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman and Hall. MR1082147. 599

Hendricks, W. and Koenker, R. (1992). "Hierarchical spline models for conditional quantiles and the demand for electricity." *Journal of the American Statistical Association*, 87(417): 58–68. 579

Huang, Y. (2016). "Quantile regression-based Bayesian semiparametric mixed-effects models for longitudinal data with non-normal, missing and mismeasured covariate." *Journal of Statistical Computation and Simulation*, 86(6): 1183–1202. MR3441563. doi: https://doi.org/10.1080/00949655.2015.1057732. 580

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): 5666–5685. MR3580932. doi: https://doi.org/10.1002/sim.7092. 580

Huang, Y., Chen, J., and Qiu, H. (2017). "Bayesian quantile regression for nonlinear mixed-effects joint models for longitudinal data in the presence of mismeasured covariate errors." *Journal of Biopharmaceutical Statistics*, 1–15. 580

Ibrahim, J. G., Lipsitz, S. R., and Chen, M. H. (1999). "Missing covariates in generalized linear models when the missing data mechanism is non-ignorable." *Journal of*

the Royal Statistical Society B, 61(1): 173–190. MR1664045. doi: https://doi.org/10.1111/1467-9868.00170. 582, 583

Ju, Y., Tang, N., and Li, X. (2018). "Bayesian local influence analysis of skew-normal spatial dynamic panel data models." *Journal of Statistical Computation and Simulation*, 88(12): 2342–2364. MR3808761. doi: https://doi.org/10.1080/00949655.2018.1462813. 580

Koenker, R. (2004). "Quantile regression for longitudinal data." *Journal of Multivariate Analysis*, 91(1): 74–89. MR2083905. doi: https://doi.org/10.1016/j.jmva.2004.05.006. 586

Koenker, R. (2005). *Quantile regression*. Cambridge University Press. 579

Kottas, A. and Krnjajić, M. (2009). "Bayesian semiparametric modelling in quantile regression." *Scandinavian Journal of Statistics*, 36(2): 297–319. MR2528986. doi: https://doi.org/10.1111/j.1467-9469.2008.00626.x. 580

Kozumi, H. and Kobayashi, G. (2011). "Gibbs sampling methods for Bayesian quantile regression." *Journal of Statistical Computation and Simulation*, 81(11): 1565–1578. MR2851270. doi: https://doi.org/10.1080/00949655.2010.496117. 581

Lancaster, T. and Sung, J. (2010). "Bayesian quantile regression methods." *Journal of Applied Econometrics*, 25(2): 287–307. MR2758636. doi: https://doi.org/10.1002/jae.1069. 579

Lee, S. Y. and Tang, N. S. (2006). "Bayesian analysis of nonlinear structural equation models with nonignorable missing data." *Psychometrika*, 71(3): 541–564. MR2272542. doi: https://doi.org/10.1007/s11336-006-1177-1. 583

Leng, C., Tran, M. N., and Nott, D. (2014). "Bayesian adaptive lasso." *Annals of the Institute of Statistical Mathematics*, 66(2): 221–244. MR3171404. doi: https://doi.org/10.1007/s10463-013-0429-6. 586

Li, Q., Xi, R., and Lin, N. (2010). "Bayesian regularized quantile regression." *Bayesian Analysis*, 1(1): 1–26. MR2719666. doi: https://doi.org/10.1214/10-BA521. 580

Liu, C. (2004). "Robit regression: a simple robust alternative to logistic and probit regression." In: A. Gelman, and X. L. Meng (Eds.), *Applied Bayesian modeling and causal inference from incomplete data perspectives* (pp. 227–238). West Sussex: Wiley. MR2138259. doi: https://doi.org/10.1002/0470090456.ch21. 599

Park, T. and Casella, G. (2008). "The Bayesian lasso." *Journal of the American Statistical Association*, 103(482): 681–686. MR2524001. doi: https://doi.org/10.1198/016214508000000337. 586

Poon, W. and Poon, Y. (1999). "Conformal normal curvature and assessment of local influence." *Journal of the Royal Statistical Society B*, 61(1): 51–61. MR1664096. doi: https://doi.org/10.1111/1467-9868.00162. 589

Pregibon, D. (1982). "Resistant fits for some commonly used logistic models with medical applications." *Biometrics*, 38: 485–498. 599

Reich, B. J., Bondell, H. D., and Wang, H. J. (2009). "Flexible Bayesian quantile regression for independent and clustered data." *Biostatistics*, 11(2): 337–352. 579

Richardson, S. and Green, P. J. (1997). "On Bayesian analysis of mixtures with an unknown number of components." *Journal of the Royal Statistical Society B*, 59(4): 731–792. MR1483213. doi: https://doi.org/10.1111/1467-9868.00095. 584

Schmid, M. and Hothorn, T. (2008). "Boosting additive models using component-wise P-splines." *Computational Statistics and Data Analysis*, 53(2): 298–311. MR2649086. doi: https://doi.org/10.1016/j.csda.2008.09.009. 599

Tang, N., Chow, S., Ibrahim, J., and Zhu, H. (2017). "Bayesian sensitivity analysis of a nonlinear dynamic factor analysis model with nonparametric prior and possible nonignorable missingness." *Psychometrika*, 82(4): 875–903. MR3736334. doi: https://doi.org/10.1007/s11336-017-9587-4. 580

Tang, N. and Zhao, H. (2014). "Bayesian analysis of nonlinear reproductive dispersion mixed models for longitudinal data with nonignorable missing covariates." *Communications in Statistics – Simulation and Computation*, 43(6): 1265–1287. MR3215778. doi: https://doi.org/10.1080/03610918.2012.732175. 584

Tanner, M. A. and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation." *Journal of the American Statistical Association*, 82(398): 528–540. MR0898357. 585

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society B*, 267–288. MR1379242. 586

Wang, Z.-Q. and Tang, N.-S. (2019). "Supplementary Material of "Bayesian Quantile regression with mixed discrete and nonignorable missing covariates"." *Bayesian Analysis*. doi: https://doi.org/10.1214/19-BA1165SUPP. 580

Wei, Y., Ma, Y., and Carroll, R. J. (2012). "Multiple imputation in quantile regression." *Biometrika*, 99(2): 423. MR2931263. doi: https://doi.org/10.1093/biomet/ass007. 580

Wei, Y. and Yang, Y. (2014). "Quantile regression with covariates missing at random." *Statistica Sinica*, 24(3): 1277–1299. MR3241288. 580

Wu, Y. and Liu, Y. (2009). "Variable selection in quantile regression." *Statistica Sinica*, 19(1): 801–817. MR2514189. 579, 586

Yang, Y. and He, X. (2012). "Bayesian empirical likelihood for quantile regression." *Annals of Statistics*, 40(4): 1102–1131. MR2985945. doi: https://doi.org/10.1214/12-AOS1005. 579

Yi, G. and He, W. (2009). "Median regression models for longitudinal data with dropouts." *Biometrics*, 65(2): 618–626. MR2751487. doi: https://doi.org/10.1111/j.1541-0420.2008.01105.x. 580

Yu, K. and Moye, R. A. (2001). "Bayesian quantile regression." *Statistics and Probability Letters*, 54(4): 437–447. MR1861390. doi: https://doi.org/10.1016/S0167-7152(01)00124-9. 579, 581

Yuan, Y. and Yin, G. (2010). "Bayesian quantile regression for longitudinal studies with nonignorable missing data." *Biometrics*, 66(1): 105–114. MR2756696. doi: https://doi.org/10.1111/j.1541-0420.2009.01269.x. 580

Zhang, H., Huang, Y., Wang, W., Chen, H., and Langlandorban, B. (2017). "Bayesian quantile regression-based partially linear mixed-effects joint models for longitudinal data with multiple features." *Statistical Methods in Medical Research*, 962280217730852. MR3580932. doi: https://doi.org/10.1002/sim.7092. 580

Zhang, Y. and Tang, N. (2017). "Bayesian local influence analysis of general estimating equations with nonignorable missing data." *Computational Statistics and Data Analysis*, 105: 184–200. MR3552196. doi: https://doi.org/10.1016/j.csda.2016.08.010. 580

Zhu, H., Ibrahim, J. G., and Tang, N. (2011). "Bayesian influence analysis: a geometric approach." *Biometrika*, 98(2): 307–323. MR2806430. doi: https://doi.org/10.1093/biomet/asr009. 580, 587, 589, 600

Zhu, H., Ibrahim, J. G., and Tang, N. (2014). "Bayesian sensitivity analysis of statistical models with missing data." *Statistica Sinica*, 24(2): 871–896. MR3235403. 589

Zou, H. (2006). "The adaptive lasso and its oracle properties." *Journal of the American Statistical Association*, 101(476): 1418–1429. MR2279469. doi: https://doi.org/10.1198/016214506000000735. 586

Acknowledgments

The authors thank the Editor, the Associate Editor and two anonymous referees for valuable comments and suggestions. This work was supported by the grants from the Key Projects of the National Natural Science Foundation of China (Grant No.: 11731101) and the National Natural Science Foundation of China (Grant No.: 11671349).