Research Article

Bayesian Adaptive Lasso for Regression Models with Nonignorable Missing Responses

Yuanying Zhao ¹ and Xingde Duan²

¹College of Mathematics and Information Science, Guiyang University, Guiyang 550 005, China

²School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550 025, China

Correspondence should be addressed to Yuanying Zhao; zhaoyuanying_@126.com

Received 20 October 2021; Revised 14 January 2022; Accepted 17 January 2022; Published 11 February 2022

Academic Editor: Niansheng Tang

Copyright © 2022 Yuanying Zhao and Xingde Duan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The main purpose of this article is to develop a Bayesian adaptive lasso procedure for analyzing linear regression models with nonignorable missing responses, in which the missingness mechanism is specified by a logistic regression model. A sampling procedure combining the Gibbs sampler and Metropolis-Hastings algorithm is employed to obtain the Bayesian estimates of the regression coefficients, shrinkage coefficients, missingness mechanism models parameters, and their standard errors. We extend the partial posterior predictive p value for goodness-of-fit statistic to investigate the plausibility of the posited model. Finally, several simulation studies and the air pollution data example are undertaken to demonstrate the newly developed methodologies.

1. Introduction

As variable selection plays an increasingly important role in modern data analysis, the least absolute shrinkage and selection operator (lasso) is appealing owing to its sparse representation. For example, see Tibshirani [1], Efron et al. [2], Zou [3], Yuan and Lin [4], Meinshausen and Bühlmann [5], etc. In the Bayesian statistical framework, Park and Casella [6] first proposed Bayesian lasso, which exploits model inference via posterior distributions. Afterwards, Guo et al. [7] studied how to select variables using Bayesian lasso in semiparametric structural equation models. Recently, Hans [8] further addressed model uncertainty and variables selection in Bayesian lasso. In particular, Bayesian adaptive lasso proposed by Leng et al. [9] permitted a treatment for variable selection with flexible penalties.

On the other hand, missing data are frequently encountered in various settings, consisting of clinical trials, surveys, and longitudinal studies. Covariates and/or responses may be missing, and statistical analysis to handle the missing data is often based on the missing data mechanism [10], such as missing not at random, also referred to as nonignorable missingness. Variable selection procedures in

the presence of missing data have received much attention in the recent literature. For instance, Ibrahim et al. [11] employed EM algorithm to propose a model selection criteria in missing data problems. After that, Garcia et al. [12] investigated the variable selection problem for regression models with missing covariate and/or response data. Recently, in the framework of semiparametric varying-coefficient partially linear models with missing response at random, a variable selection procedure presented by Zhao and Xue [13] can be used to simultaneously select significant variables in nonparametric and parametric model, among others. However, to the best of our knowledge, the study of missing data based on Bayesian adaptive Lasso et al. [9] has not been proposed at present. Therefore, our motivation is to construct a Bayesian method to simultaneously select variables and estimate parameters in linear regression models with nonignorable missing responses under the framework of Bayesian adaptive lasso. In our work, missing data mechanism is specified by a logistic regression model, and a MCMC procedure is proposed to obtain the Bayesian estimates of the regression coefficients, shrinkage coefficients, missingness mechanism models parameters, and their standard errors. Contributions of our work mainly include

(1) tracking with the computationally intractable integral problem in posterior analysis, especially when nonignorable missing data raises new challenges for variable selection in our considered models; (2) the sampling-based approach could give rise to reliable statistical inferences and does not depend on asymptotic theory.

The structure of the article is arranged as follows.

We present a hierarchical Bayesian lasso model in the framework of linear regression models with nonignorable missing responses in Section 2. In Section 3, a Bayesian procedure is proposed and developed by combining the Gibbs sampler [14] and the Metropolis-Hastings (M-H) algorithm [15, 16] to obtain statistical inference for simultaneous variable selection and parameter estimation. In particular, the partial posterior predictive p -value for goodness-of-fit statistic is also presented to investigate the plausibility of the posited model in Section 3. Several simulation studies and the air pollution data example are employed to demonstrate the preceding proposed Bayesian methodologies in Section 4.

The article is ended up with a discussion in Section 5.

2. Model and Notations

We take into account the linear regression model problem $\mathbf{y} = \mu \mathbf{1}_n + \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, in which $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ is an $n \times 1$ vector of responses, $\mathbf{x}$ is an $n \times p$ matrix of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ regression parameter, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ is an $n \times 1$ vector of normal errors with mean 0_n and covariance $\sigma^2 \mathbf{I}_n$. Here, $\mathbf{1}_n$ (or 0_n) is an $n \times 1$ vector whose each element is to set 1 (or 0), $\mathbf{I}_n$ represents the $n \times n$ identity matrix, and the superscript $'$ denotes the transposition of a vector (or matrix). Without loss of generality, we just consider the case that μ is 0 and may be deleted from the above proposed model in this article since $\mathbf{y}$ and $\mathbf{x}$ can be centered. Similar to Leng et al. [9], the following adaptive lasso estimator is used to obtain statistical inference for simultaneous parameter estimation and variable selection

$$\min (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) + \sum_{j=1}^p \lambda_j |\beta_j|. \quad (1)$$

Here, flexible penalty parameters are used for different regression parameters. The following hierarchical Bayesian lasso model is considered in this article:

$$\begin{aligned} \mathbf{y} | \mathbf{x}, \boldsymbol{\beta}, \sigma^2 &\sim N_n(\mathbf{x}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \\ \boldsymbol{\beta} | \sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim N_p(0_p, \sigma^2 \mathbf{D}_\tau), \\ \mathbf{D}_\tau &= \text{diag}(\tau_1^2, \dots, \tau_p^2), \end{aligned} \quad (2)$$

with the following priors on σ^2 and $\tau^2 = (\tau_1^2, \dots, \tau_p^2)'$:

$$\sigma^2, \tau_1^2, \dots, \tau_p^2 \sim \pi(\sigma^2) \prod_{j=1}^p \frac{\lambda_j^2}{2} \exp \left\{ -\frac{\lambda_j^2 \tau_j^2}{2} \right\}. \quad (3)$$

Park and Casella [6] proposed to employ the improper prior $\pi(\sigma^2) \propto 1/\sigma^2$ to model the error variance, where $N_p(\cdot, \cdot)$ denotes the p -dimensional normal distribution.

In this article, we focus on the case that the responses $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ are incompletely observed, which may be subject to nonignorable missingness, while covariates $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$ are completely observed, where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$. For $\mathbf{y} = (y_1, y_2, \dots, y_n)'$, we define a vector $\mathbf{r} = (r_1, r_2, \dots, r_n)'$ of missing indicators such that $r_i = 0$ if it is observed and $r_i = 1$ if y_i is missing. Therefore, $\mathbf{y}$ can be expressed by $\mathbf{y} = \{\mathbf{y}_o, \mathbf{y}_m\}$, where $\mathbf{y}_o$ and $\mathbf{y}_m$ denote the set of the observed data and the missing values associated with the responses, respectively. Let $p(r_i | y_i, \mathbf{x}_i, \boldsymbol{\varphi})$ be the conditional probability distribution of r_i given $y_i, \mathbf{x}_i$ and unknown parameter $\boldsymbol{\varphi}$, and then

$$\begin{aligned} p(r_i | y_i, \mathbf{x}_i, \boldsymbol{\varphi}) &= \{p(r_i = 1 | y_i, \mathbf{x}_i, \boldsymbol{\varphi})\}^{r_i} \\ &\quad \{1 - p(r_i = 1 | y_i, \mathbf{x}_i, \boldsymbol{\varphi})\}^{1-r_i}. \end{aligned} \quad (4)$$

Here, $p(r_i = 1 | y_i, \mathbf{x}_i, \boldsymbol{\varphi})$ can be specified by the following logistic regression model:

$$\begin{aligned} \text{logit}\{p(r_i = 1 | y_i, \mathbf{x}_i, \boldsymbol{\varphi})\} &= \varphi_0 + \varphi_1 y_i + \varphi_{21} x_{i1} + \dots \\ &\quad + \varphi_{2p} x_{ip} \triangleq \boldsymbol{\varphi}' \boldsymbol{\omega}_i, \end{aligned} \quad (5)$$

in which $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \varphi_{21}, \dots, \varphi_{2p})'$ and $\boldsymbol{\omega}_i = (1, y_i, x_{i1}, \dots, x_{ip})'$.

Let $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2, \tau^2, \lambda^2\}$ that contains all unknown parameter in models (1)–(3), in which $\tau^2 = (\tau_1^2, \dots, \tau_p^2)'$ and $\lambda^2 = (\lambda_1^2, \dots, \lambda_p^2)'$. The main objective of this article is to develop a Bayesian procedure to analyze the above given model under the missing data indicators $\mathbf{r}$ and the observed data $\{\mathbf{y}_o, \mathbf{x}\}$. To make Bayesian statistical inferences on the above given model, random observations are needed to be generated from the posterior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$. For the conditional distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ given the missing data indicators $\mathbf{r}$ and the observed data $\{\mathbf{y}_o, \mathbf{x}\}$, it follows from models (1)–(5) that $p(\boldsymbol{\theta}, \boldsymbol{\varphi} | \mathbf{y}_o, \mathbf{x}, \mathbf{r}) \propto$

$$p(\mathbf{y}_o, \mathbf{x}, \mathbf{r} | \boldsymbol{\theta}, \boldsymbol{\varphi}) p(\boldsymbol{\theta}, \boldsymbol{\varphi}) \propto \left\{ \int \prod_{i=1}^n p(y_i | \mathbf{x}_i, \boldsymbol{\theta}) p(r_i | y_i, \mathbf{x}_i, \boldsymbol{\varphi}) d\mathbf{y}_m \right\} p(\boldsymbol{\theta}, \boldsymbol{\varphi}), \quad (6)$$

where $p(\theta, \varphi)$ represents the joint prior probability density for θ and φ . Obtaining a closed form of integral (6) is rather difficult due to the missing mechanism model involved.

3. Bayesian Analysis of the Models

For developing an effective Bayesian approach to analyze the above given model and reducing computational burden, we augment the indicators $\mathbf{r}$ and the observed data $\{\mathbf{y}_o, \mathbf{x}\}$ with the missing values $\mathbf{y}_m$ to produce full data set $\{\mathbf{y}_o, \mathbf{x}, \mathbf{y}_m, \mathbf{r}\} = \{\mathbf{y}, \mathbf{x}, \mathbf{r}\}$ in the posterior analysis on the basis of the data augmentation idea [17]. Clearly, the posterior density $p(\theta, \varphi, \mathbf{y}_m | \mathbf{y}_o, \mathbf{x}, \mathbf{r})$ is easier to handle than $p(\theta, \varphi | \mathbf{y}_o, \mathbf{x}, \mathbf{r})$.

However, generating observations from $p(\theta, \varphi, \mathbf{y}_m | \mathbf{y}_o, \mathbf{x}, \mathbf{r})$ is still rather difficult due to the complexity of the considered model.

The Gibbs sampler [14], one of the most useful tools in statistical calculation, is employed to simulate a sequence of samples from the joint posterior distribution $p(\theta, \varphi, \mathbf{y}_m | \mathbf{y}_o, \mathbf{x}, \mathbf{r})$ and then make Bayesian statistical inferences on the above given model.

The key to this algorithm is that random samples $\{\theta, \varphi, \mathbf{y}_m\}$ are iteratively simulated from the following three conditional distribution functions: $p(\theta | \mathbf{y}_o, \mathbf{y}_m, \mathbf{x}, \varphi, \mathbf{r})$, $p(\varphi | \mathbf{y}_o, \mathbf{y}_m, \mathbf{x}, \theta, \mathbf{r})$, and $p(\mathbf{y}_m | \mathbf{y}_o, \mathbf{x}, \theta, \varphi, \mathbf{r})$. Note that $p(\theta | \mathbf{y}_o, \mathbf{y}_m, \mathbf{x}, \varphi, \mathbf{r}) = p(\theta | \mathbf{y}, \mathbf{x})$ and $p(\varphi | \mathbf{y}_o, \mathbf{y}_m, \mathbf{x}, \theta, \mathbf{r}) = p(\varphi | \mathbf{y}, \mathbf{x}, \theta, \mathbf{r})$ in our considered model.

3.1. Gibbs Sampling and Conditional Distributions. Based on the above discussion, the Gibbs sampler [14] is implemented as follows:

Step 1. Specifying the initial values for parameters θ, φ and $\mathbf{y}_m$, let $\theta^{(0)} = \{\beta^{(0)}, \sigma^{2(0)}, \tau^{2(0)}, \lambda^{2(0)}\}$, $(\varphi^{(0)} = (\varphi_0^{(0)}, \varphi_1^{(0)}, \varphi_{21}^{(0)}, \dots, \varphi_{2p}^{(0)})')$ and $\mathbf{y}_m = \mathbf{y}_m^{(0)}$.

Step 2. Based on $\theta^{(t)} = \{\beta^{(t)}, \sigma^{2(t)}, \tau^{2(t)}, \lambda^{2(t)}\}$, and $\mathbf{y}^{(t)} = \{\mathbf{y}_o, \mathbf{y}_m^{(t)}\}$, generating sample $\theta^{(t+1)} = \{\beta^{(t+1)}, \sigma^{2(t+1)}, \tau^{2(t+1)}, \lambda^{2(t+1)}\}$ as follows:

Generating sample $\beta^{(t+1)}$ from the following distribution:

$$p(\beta | \mathbf{y}^{(t)}, \mathbf{x}, \sigma^{2(t)}, \tau^{2(t)}) \sim N_p(\boldsymbol{\mu}^{(t)}, \sigma^{2(t)} \mathbf{A}^{(t)-1}), \quad (7)$$

where $\mathbf{A}^{(t)} = \mathbf{x}'\mathbf{x} + \mathbf{D}_\tau^{-1}$ and $\boldsymbol{\mu}^{(t)} = \mathbf{A}^{(t)-1}\mathbf{x}'\mathbf{y}^{(t)}$ in which $\mathbf{D}_\tau = \text{diag}(\tau_1^{2(t)}, \dots, \tau_p^{2(t)})$.

Generating sample $\sigma^{2(t+1)}$ from the following distribution:

$$p\left(\frac{1}{\sigma^2} | \mathbf{y}^{(t)}, \mathbf{x}, \beta^{(t+1)}, \tau^{2(t)}\right) \sim \Gamma(sh, sc^{(t)}), \quad (8)$$

where $\Gamma(a_1, a_2)$ represents Gamma distribution with shape parameter a_1 and scale parameter a_2 , in which $sh = n + p - 1/2$, $sc^{(t)} = (\mathbf{y}^{(t)} - \mathbf{x}\beta^{(t+1)})'(\mathbf{y}^{(t)} - \mathbf{x}\beta^{(t+1)})/2 + \beta^{(t+1)'}\mathbf{D}_\tau^{-1}\beta^{(t+1)}/2$.

Generating sample $\tau^{2(t+1)}$ from the following distribution:

$$p\left(\frac{1}{\tau_j^2} | \beta^{(t+1)}, \sigma^{2(t+1)}, \lambda_j^{2(t)}\right) \sim IG(\mu_j^*, \lambda_j^*). \quad (9)$$

for $j = 1, 2, \dots, p$,

where $IG(a_3, a_4)$ represents inverse Gaussian distribution with parameters a_3 and a_4 , in which $\mu_j^* = \lambda_j^{(t+1)}\sigma^{(t+1)}/|\beta_j^{(t+1)}|$ and $\lambda_j^* = \lambda_j^{2(t)}$.

Sampling $\lambda^{2(t+1)}$:

There are two methods to update parameter $\lambda^{2(t+1)}$: empirical Bayesian method and hierarchical model method; see more details in Leng et al. [9].

Step 3. Based on $\mathbf{y}^{(t)} = \{\mathbf{y}_o, \mathbf{y}_m^{(t)}\}$,

Generating sample $\varphi^{(t+1)} = (\varphi_0^{(t+1)}, \varphi_1^{(t+1)}, \varphi_{21}^{(t+1)}, \dots, \varphi_{2p}^{(t+1)})'$ from the following distribution:

$$p(\boldsymbol{\varphi} | \mathbf{y}^{(t)}, \mathbf{x}, \mathbf{r}) \propto p(\mathbf{r} | \mathbf{y}^{(t)}, \mathbf{x}, \boldsymbol{\varphi}) p(\boldsymbol{\varphi}) \propto \frac{\exp\{\sum_{i=1}^n r_i \varphi' \boldsymbol{\omega}_i - 1/2(\boldsymbol{\varphi} - \boldsymbol{\varphi}^0)' V^{-1}(\boldsymbol{\varphi} - \boldsymbol{\varphi}^0)\}}{\prod_{i=1}^n (1 + \exp(\varphi' \boldsymbol{\omega}_i))}, \quad (10)$$

where $\boldsymbol{\omega}_i = (1, y_i^{(t)}, \mathbf{x}_i')'$, $p(\boldsymbol{\varphi}) \sim N_{p+2}(\boldsymbol{\varphi}^0, V)$ is the prior distribution of parameter $\boldsymbol{\varphi}$.

Step 4. Based on $\theta^{(t+1)}$, $\varphi^{(t+1)}$ and $\mathbf{y}^{(t)} = \{\mathbf{y}_o, \mathbf{y}_m^{(t)}\}$, generating sample $\mathbf{y}_m^{(t+1)}$ from the following distribution:

$$p(y_{im} | \mathbf{x}_i, \boldsymbol{\theta}^{(t+1)}, \boldsymbol{\varphi}^{(t+1)}) \propto \exp\left\{-\frac{1}{2\sigma^{2(t+1)}}(y_{im} - \mathbf{x}_i' \boldsymbol{\beta}^{(t+1)})^2 + (\boldsymbol{\varphi}^{(t+1)})^T \boldsymbol{\omega}_i - \log(1 + \exp((\boldsymbol{\varphi}^{(t+1)})' \boldsymbol{\omega}_i))\right\}, \quad (11)$$

where $\boldsymbol{\omega}_i = (1, y_i^{(t)}, \mathbf{x}_i')'$.

Step 5. Repeating steps 2 – 4.

Followed by Geman and Geman [14], the sequence of observation $\{\theta^{(t)}, \varphi^{(t)}, \mathbf{y}_m^{(t)}: t = 1, 2, \dots, T\}$ can be treated as a sample simulated from the joint posterior function $p(\theta, \varphi, \mathbf{y}_m | \mathbf{y}_o, \mathbf{x}, \mathbf{r})$. In practice, the estimated potential scale

reduction (EPSR) values [18] can be employed to monitor the convergence of the algorithm.

It is easy to find that conditional distributions from (7)–(9) are some standard and familiar distributions. Therefore, generating samples from these conditional distributions are direct. However, those corresponding conditional distributions given in (10) and (11) are complicated

and nonstandard distributions. As a result, to directly generate samples from those conditional distributions is rather difficult. The well-known M-H algorithm [15, 16] is one of the most effective ways to generate samples from those corresponding conditional distributions with the help of proposal distributions. Similar to Zhao and Tang [19] and Xu et al. [20], the following proposal distributions $N_{p+2}(0, \sigma_\varphi^2 \Omega_\varphi)$ and $N_1(0, \sigma_y^2 \Omega_y)$ are employed for sampling φ and y_{im} , respectively, where $\Omega_\varphi^{-1} = 1/4 \sum_{i=1}^n \omega_i \omega_i' + \mathbf{V}^{-1}$ and $\Omega_y^{-1} = \{\exp(\varphi' \omega_i) / [1 + \exp(\varphi' \omega_i)]^2 \varphi_1^2 + 1/\sigma^2\}_{y_i=0}$. Specifically, the M-H algorithm for sampling observation φ and y_{im} is implemented by the following process: at the $(t+1)$ th iteration with the current value $\varphi^{(t)}$ and $y_{im}^{(t)}$, new candidates φ^* and y_i^* are generated from $N_{p+2}(\varphi^{(t)}, \sigma_\varphi^2 \Omega_\varphi)$ and $N_1(y_{im}^{(t)}, \sigma_y^2 \Omega_y)$, respectively. They are accepted with probabilities $\min\{1, p(\varphi^*|\mathbf{y}, \mathbf{x}, \mathbf{r})/p(\varphi^{(t)}|\mathbf{y}, \mathbf{x}, \mathbf{r})\}$ and $\min\{1 + p(y_i^*|\mathbf{x}_i, \theta, \varphi)/p(y_{im}^{(t)}|\mathbf{x}_i, \theta, \varphi)\}$, respectively. σ_φ^2 and σ_y^2 are chosen such that the average acceptance rate is between 0.25 and 0.45 [21].

3.2. Bayesian Estimates and Goodness-of-Fit Statistic. The joint Bayesian estimates of parameters θ, φ and $\mathbf{y}_m$ are expressed by

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^T \theta^{(t)}, \hat{\varphi} = \frac{1}{T} \sum_{t=1}^T \varphi^{(t)}, \hat{\mathbf{y}}_m = \frac{1}{T} \sum_{t=1}^T \mathbf{y}_m^{(t)}, \quad (12)$$

where $\{(\theta^{(t)}, \varphi^{(t)}, \mathbf{y}_m^{(t)}): t = 1, 2, \dots, T\}$ are the samples that are obtained from the joint posterior distribution $p(\theta, \varphi, \mathbf{y}_m|\mathbf{y}_o, \mathbf{x}, \mathbf{r})$ via the previous Gibbs sampler procedure. Following Geyer [22], these joint Bayesian estimates and sample covariance matrices are consistent with their corresponding posterior expectations and posterior covariance matrices. Therefore, we can obtain the standard errors by the diagonal elements of the sample covariance matrices.

The partial posterior predictive (PPP) p -value proposed by Bayarri and Berger [23] is a goodness-of-fit statistic to investigate the plausibility of the posited model in Bayesian statistical framework; see more details in Lee and Tang [24], Tang and Zhao [25]. Similar to Lee and Tang [24] and Tang and Zhao [25], we can define the PPP p -value for our considered model as

$$ppp_B = \int \Pr\{D(\mathbf{y}_o|\theta, \varphi, \mathbf{y}_m, \mathbf{x}, \mathbf{r}) \geq D(\mathbf{y}_o^{\text{obs}}|\theta, \varphi, \mathbf{y}_m, \mathbf{x}, \mathbf{r})\} \cdot p^*(\theta, \varphi, \mathbf{y}_m) d\theta d\varphi d\mathbf{y}_m, \quad (13)$$

in which $\mathbf{y}_o^{\text{obs}}$ is the observed value of the responses $\mathbf{y}_o$ and

$$p^*(\theta, \varphi, \mathbf{y}_m) \propto \frac{p(\mathbf{y}_o^{\text{obs}}|\mathbf{y}_m, \mathbf{x}, \mathbf{r}|\theta, \varphi) p(\theta, \varphi)}{p(D^{\text{obs}}|\theta, \varphi, \mathbf{y}_m)}, \quad (14)$$

$D(\cdot|\cdot)$ in (13) denotes a discrepancy variable, and D^{obs} represents the observed value of $D(\cdot|\cdot)$ on the basis of $\mathbf{y}_o^{\text{obs}}$. The following discrepancy variable can be specified in our problem:

$$D(\mathbf{y}_o|\theta, \varphi, \mathbf{y}_m, \mathbf{x}, \mathbf{r}) = \sum_{i=1}^n d_i, \quad (15)$$

in which

$$d_i = \begin{cases} (y_{oi} - v_{oi}^*) V_{oi}^{-1} (y_{oi} - v_{oi}^*), & \text{if } r_i = 0; \\ 0, & \text{if } r_i = 1. \end{cases} \quad (16)$$

where v_{oi}^* and V_{oi} are conditional expectation and conditional variance of y_{oi} given $\theta, \varphi, \mathbf{y}_m, \mathbf{x}$ and $\mathbf{r}$. It is easily seen that

$$p^*(\theta, \varphi, \mathbf{y}_m) \propto p(\mathbf{y}_o^{\text{obs}}|\mathbf{y}_m, \mathbf{x}, \mathbf{r}|\theta, \varphi) p(\theta, \varphi) \propto \left\{ \prod_{i=1}^n p(y_i|\mathbf{x}_i, \theta) p(r_i|y_i, \mathbf{x}_i, \theta, \varphi) \right\} p(\theta, \varphi). \quad (17)$$

To reduce computational burden of the PPP p -value in (13), the Monte Carlo integral is employed to approximate ppp_B . Let $\{(\theta^{(t)}, \varphi^{(t)}, \mathbf{y}_m^{(t)}): t = 1, 2, \dots, T\}$ be the samples generated from the distribution $p^*(\theta, \varphi, \mathbf{y}_m)$ via the previous Gibbs sampler procedure, and let $\{\mathbf{y}_o^{(t)}: t = 1, \dots, T\}$ be the samplers simulated from $p(\mathbf{y}_o|\mathbf{y}_m, \mathbf{x}, \theta, \varphi, \mathbf{r})$. Then, the PPP p -value in (13) is approximated by

$$\begin{aligned} \widehat{ppp}_B &= \frac{1}{T} \sum_{t=1}^T I\{D(\mathbf{y}_o^{(t)}|\theta^{(t)}, \varphi^{(t)}, \mathbf{y}_m^{(t)}, \mathbf{x}, \mathbf{r}) \\ &\geq D(\mathbf{y}_o^{\text{obs}}|\theta^{(t)}, \varphi^{(t)}, \mathbf{y}_m^{(t)}, \mathbf{x}, \mathbf{r})\}, \end{aligned} \quad (18)$$

where $I(\cdot)$ is an indicator function. According to [24, 25], if $\widehat{ppp}_B$ is far away from 0.5, then we reject the posited model.

4. Numerical Examples

Several simulation studies and the air pollution data example are employed to demonstrate our proposed methodologies in this section.

4.1. Example 1. The first simulation study is employed to illustrate the performance of the Bayesian estimates with respective to the sparse, dense, and very sparse cases,

respectively. Specifically, we generate observations y_i from the following model:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + \varepsilon_i. \quad (19)$$

For $i = 1, 2, \dots, n$, here, the true values for parameter $\boldsymbol{\beta}$ are given as follows:

Case i (moderate case): $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)^T$;

Case ii (dense case): $\boldsymbol{\beta} = (1, 1, 1, 1, 1, 1, 1, 1)^T$;

Case iii (sparse case): $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0)^T$.

For $i = 1, 2, \dots, n$, $j, k = 1, 2, \dots, 8$, it is assumed that x_{ij} follows standard normal distribution $N(0, 1)$ marginally and the correlation between x_{ij} and x_{ik} is $0.5^{|j-k|}$, and ε_i is independent and identically distributed as $N_1(0, \sigma^2)$, and the true value of σ^2 is set to 1 in all simulation studies. Missing data are generated from the logistic regression model

$$M_0: \text{logit}\{p(r_i = 1|y_i, \mathbf{x}_i, \boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 y_i + \varphi_{21} x_{i1} + \cdots + \varphi_{28} x_{i8}, \quad (20)$$

with true values $\varphi_0 = -1.0$, $\varphi_1 = \varphi_{21} = \cdots = \varphi_{28} = 0.3$. The missing data are generated as follows: (a) simulate the full data set $\{y_i: i = 1, 2, \dots, n\}$ from model (19); (b) identify whether the observation y_i is missing or not via the logistic model (20) with the true values for $\varphi_0, \varphi_1, \varphi_{21}, \dots, \varphi_{28}$. More specifically, a random number κ is obtained from the uniform distribution $U(0, 1)$; the observation y_i is missing if $\kappa \leq \text{pr}(r_i = 1|y_i, \mathbf{x}_i, \boldsymbol{\varphi}) = 1/1 + \exp\{-\varphi_0 - \varphi_1 y_i - \varphi_{21} x_{i1} - \cdots - \varphi_{28} x_{i8}\}$.

For each of the above generated data sets under various model assumptions and parameter settings, our proposed methodologies combining the Gibbs sampler together with the M-H algorithm are employed to calculate the Bayesian estimates of unknown parameters on the basis of 100 replications for $n = 200$, and the average proportion of missing data simulated are roughly 0.375. In each replication, we collected 10 000 observations after 10 000 iterations in producing Bayesian results. In M-H algorithm, we take $\sigma_y^2 = 4.0$ and $\sigma_\varphi^2 = 1.2$ obtaining average acceptance rates 29.6% and 27.5%, respectively. The corresponding results of Bayesian estimates are reported in Tables 1–3, in which ‘True’ represents the true value of parameter; ‘Mean,’ ‘SD,’ and ‘Median’ denote the average, standard deviation, and median values of the Bayesian estimates on the basis of 100 data sets, respectively; ‘5%’ and ‘95%’ are the 5% and 95% quantiles of the Bayesian estimates on the basis of 100 replications; ‘RMSE’ is the root mean square error between the Bayesian estimates on the basis of 100 data sets and its true value. Besides, to summarize the variable selection results, in Table 4, we calculate the proportion that parameter $\boldsymbol{\beta}$ was identified to be zero in 100 replications in terms of the criterion that a parameter is identified to be 0 if its 95% confidence interval contains zero. Inspections of Tables 1–4 indicate that (1) the Bayesian estimates of unknown parameters for $\boldsymbol{\beta}$, σ and $\boldsymbol{\varphi}$ are quite close to their true values under moderate case, dense case, and sparse case, which illustrates that our proposed Bayesian approach is quite effective; (2) the differences between ‘SD’ and ‘RMSE’

TABLE 1: Bayesian estimates for case i in the first simulation study.

Parameter	True	Mean	Median	5%	95%	SD	RMSE
(i) Empirical Bayesian scenario							
β_1	3.0	2.997	2.991	2.787	3.147	0.104	0.104
β_2	1.5	1.476	1.473	1.282	1.644	0.113	0.112
β_3	0.0	0.010	0.001	-0.112	0.130	0.067	0.067
β_4	0.0	0.001	-0.000	-0.090	0.097	0.062	0.062
β_5	2.0	1.992	1.981	1.836	2.180	0.111	0.111
β_6	0.0	0.007	0.000	-0.111	0.093	0.069	0.069
β_7	0.0	0.001	-0.000	-0.106	0.010	0.063	0.062
β_8	0.0	-0.007	0.000	-0.134	0.045	0.050	0.050
σ	1.0	0.999	0.998	0.887	1.089	0.064	0.063
φ_0	-1.0	-1.091	-1.102	-1.348	-0.826	0.169	0.191
φ_1	0.3	0.368	0.353	0.113	0.631	0.161	0.174
φ_{21}	0.3	0.265	0.244	-0.520	0.958	0.456	0.456
φ_{22}	0.3	0.337	0.327	-0.329	0.922	0.361	0.361
φ_{23}	0.3	0.362	0.377	-0.122	0.768	0.274	0.280
φ_{24}	0.3	0.329	0.294	-0.193	0.811	0.302	0.302
φ_{25}	0.3	0.317	0.269	-0.256	0.833	0.338	0.338
φ_{26}	0.3	0.362	0.373	-0.252	0.901	0.305	0.310
φ_{27}	0.3	0.340	0.316	-0.167	0.803	0.294	0.295
φ_{28}	0.3	0.345	0.342	-0.098	0.749	0.272	0.274
(ii) Hierarchical model scenario							
β_1	3.0	3.005	2.987	2.829	3.149	0.111	0.110
β_2	1.5	1.500	1.511	1.319	1.646	0.113	0.113
β_3	0.0	0.007	0.001	-0.059	0.066	0.047	0.048
β_4	0.0	-0.004	0.001	-0.119	0.067	0.069	0.069
β_5	2.0	2.003	1.996	1.834	2.157	0.101	0.099
β_6	0.0	-0.005	-0.002	-0.075	0.058	0.041	0.041
β_7	0.0	-0.003	-0.004	-0.077	0.061	0.040	0.039
β_8	0.0	-0.003	-0.003	-0.049	0.039	0.032	0.032
σ	1.0	1.019	1.020	0.908	1.108	0.063	0.066
φ_0	-1.0	-1.091	-1.102	-1.358	-0.853	0.161	0.184
φ_1	0.3	0.357	0.357	0.053	0.610	0.159	0.168
φ_{21}	0.3	0.288	0.257	-0.630	0.946	0.458	0.456
φ_{22}	0.3	0.281	0.266	-0.224	0.917	0.347	0.346
φ_{23}	0.3	0.384	0.360	-0.184	0.851	0.306	0.316
φ_{24}	0.3	0.360	0.337	-0.212	0.780	0.296	0.301
φ_{25}	0.3	0.299	0.348	-0.363	0.777	0.332	0.330
φ_{26}	0.3	0.328	0.308	-0.205	0.782	0.286	0.286
φ_{27}	0.3	0.384	0.417	-0.126	0.799	0.284	0.295
φ_{28}	0.3	0.299	0.311	-0.087	0.644	0.224	0.223

are very small, which shows that the Bayesian estimates are rather reliable and robust to our considered cases. (3) Our proposed methods could identify the correct models in most cases, because the proportion values in Table 4 corresponding to the important covariates are all zero, and the proportion values corresponding to unimportant covariates are more than 90%. (4) It is interesting to observe that the hierarchical model approach tends to obtain better results than empirical Bayesian approach in terms of variable selection, but the differences are minor, especially in the very dense case (e.g., case ii). To sum up, the above findings indicate that our preceding proposed Bayesian procedures could well recover the true information in models.

4.2. Example 2. In order to investigate the sensitivity to proportion of missing data, we conduct the second simulation study in this part. Simulated data sets are generated by

TABLE 2: Bayesian estimates for case ii in the first simulation study.

Parameter	True	Mean	Median	5%	95%	SD	RMSE
(i) Empirical Bayesian scenario							
β_1	1.0	1.014	1.023	0.815	1.163	0.099	0.100
β_2	1.0	0.994	0.997	0.783	1.185	0.129	0.128
β_3	1.0	0.983	0.974	0.788	1.157	0.109	0.110
β_4	1.0	1.015	1.020	0.819	1.180	0.115	0.116
β_5	1.0	1.004	1.013	0.836	1.156	0.105	0.105
β_6	1.0	0.992	0.986	0.788	1.164	0.114	0.113
β_7	1.0	0.991	0.979	0.811	1.175	0.116	0.115
β_8	1.0	0.989	0.983	0.829	1.125	0.096	0.096
σ	1.0	1.010	1.006	0.915	1.131	0.065	0.065
φ_0	-1.0	-1.092	-1.083	-1.395	-0.841	0.172	0.195
φ_1	0.3	0.351	0.345	-0.045	0.692	0.227	0.231
φ_{21}	0.3	0.317	0.304	-0.163	0.765	0.317	0.316
φ_{22}	0.3	0.331	0.278	-0.185	0.919	0.334	0.334
φ_{23}	0.3	0.371	0.365	-0.173	0.892	0.338	0.344
φ_{24}	0.3	0.310	0.286	-0.375	0.996	0.388	0.387
φ_{25}	0.3	0.332	0.302	-0.220	0.831	0.338	0.338
φ_{26}	0.3	0.316	0.309	-0.250	0.869	0.362	0.361
φ_{27}	0.3	0.363	0.329	-0.310	0.904	0.394	0.397
φ_{28}	0.3	0.318	0.286	-0.209	0.747	0.327	0.326
(ii) Hierarchical model scenario							
β_1	1.0	0.999	1.000	0.816	1.163	0.112	0.112
β_2	1.0	0.979	0.963	0.798	1.192	0.124	0.125
β_3	1.0	1.017	0.999	0.842	1.224	0.125	0.125
β_4	1.0	1.001	0.982	0.786	1.224	0.129	0.128
β_5	1.0	0.987	1.004	0.767	1.143	0.122	0.122
β_6	1.0	0.990	0.973	0.761	1.237	0.139	0.139
β_7	1.0	0.993	0.991	0.814	1.165	0.114	0.114
β_8	1.0	0.996	1.002	0.807	1.165	0.111	0.110
σ	1.0	1.019	1.021	0.915	1.129	0.067	0.069
φ_0	-1.0	-1.087	-1.094	-1.356	-0.843	0.167	0.188
φ_1	0.3	0.375	0.343	0.051	0.712	0.209	0.221
φ_{21}	0.3	0.361	0.389	-0.168	0.739	0.295	0.300
φ_{22}	0.3	0.325	0.340	-0.482	0.849	0.387	0.386
φ_{23}	0.3	0.325	0.360	-0.287	0.788	0.332	0.332
φ_{24}	0.3	0.323	0.334	-0.304	0.903	0.386	0.385
φ_{25}	0.3	0.292	0.279	-0.342	0.847	0.360	0.359
φ_{26}	0.3	0.295	0.282	-0.277	0.903	0.349	0.347
φ_{27}	0.3	0.340	0.356	-0.256	0.854	0.352	0.353
φ_{28}	0.3	0.311	0.302	-0.240	0.780	0.306	0.304

using the same setup as specified in the first simulation study except that true value of φ is set as $\varphi_0 = -2.5$, $\varphi_1 = \varphi_{21} = \dots = \varphi_{28} = 0.3$. We fit these simulated data sets under cases i-iii via the preceding developed MCMC algorithm and only present results in Table 5 on the basis of case i under 100 replications to save space, where the average proportion of missing data generated in the second simulation study is about 0.208. Besides, we further calculate sums of RMSE under case i based on these two simulation studies, and the results are 3.365 and 3.237 in this simulation study, compared with 3.681 and 3.522 in the first simulation study. Moreover, the median of mean absolute deviations (MMAD) is calculated for 100 replications, i.e., median $\{1/n \sum_{i=1}^n |\hat{\mathbf{x}}_i \hat{\beta} - \mathbf{x}_i \beta^{true}|\}$, in which $\hat{\beta}$ and β^{true} represent the Bayesian estimate and true value of parameter β .

The related results of MMAD for the first (second) simulation are 0.143 7 (0.129 3) and 0.134 8 (0.122 1) based

on Empirical Bayesian method and Hierarchical model method, respectively. Meanwhile, the box plots for mean absolute deviations (MAD) are shown in Figure 1, where “1” and “2” denote case i in the first and second simulation studies, respectively. All these results indicate that cases with the smaller proportion of missing data in this simulation study generally perform better than cases with the larger proportion of missing data in the first simulation study.

4.3. Example 3. To address the influence of the missingness mechanism model to the Bayesian adaptive lasso, we conduct the third simulation in this example. The full data set is created by employing the same setting specified under case i in the first simulation study. But, we generate the missing data via the following 3 types of missingness mechanisms.

Type i: logistic regression model that is different from M_0 specified in (20):

TABLE 3: Bayesian estimates for case iii in the first simulation study.

Parameter	True	Mean	Median	5%	95%	SD	RMSE
(i) Empirical Bayesian scenario							
β_1	5.0	5.003	4.992	4.826	5.185	0.111	0.111
β_2	0.0	-0.011	-0.000	-0.153	0.046	0.068	0.068
β_3	0.0	0.003	-0.001	-0.060	0.053	0.054	0.054
β_4	0.0	0.012	-0.001	-0.193	0.085	0.076	0.076
β_5	0.0	0.008	0.001	-0.078	0.091	0.065	0.066
β_6	0.0	0.007	0.001	-0.083	0.082	0.111	0.111
β_7	0.0	0.004	0.001	-0.107	0.080	0.053	0.052
β_8	0.0	0.000	0.000	-0.119	0.078	0.064	0.063
σ	1.0	0.991	0.984	0.885	1.080	0.058	0.058
φ_0	-1.0	-1.065	-1.071	-1.310	-0.798	0.163	0.175
φ_1	0.3	0.359	0.361	0.098	0.529	0.123	0.136
φ_{21}	0.3	0.273	0.243	-0.442	0.958	0.443	0.442
φ_{22}	0.3	0.323	0.288	-0.073	0.746	0.285	0.285
φ_{23}	0.3	0.367	0.324	0.006	0.745	0.244	0.252
φ_{24}	0.3	0.318	0.307	-0.171	0.710	0.262	0.262
φ_{25}	0.3	0.346	0.296	-0.109	0.822	0.296	0.298
φ_{26}	0.3	0.322	0.339	-0.187	0.783	0.297	0.296
φ_{27}	0.3	0.326	0.300	-0.023	0.668	0.228	0.229
φ_{28}	0.3	0.342	0.341	-0.038	0.722	0.234	0.237
(ii) Hierarchical model scenario							
β_1	5.0	5.009	5.006	4.875	5.163	0.091	0.091
β_2	0.0	-0.000	0.001	-0.058	0.037	0.038	0.038
β_3	0.0	-0.001	-0.002	-0.046	0.035	0.031	0.031
β_4	0.0	0.004	-0.000	-0.042	0.048	0.029	0.029
β_5	0.0	0.002	0.002	-0.050	0.063	0.036	0.036
β_6	0.0	-0.001	0.000	-0.058	0.037	0.038	0.038
β_7	0.0	0.006	0.005	-0.043	0.055	0.030	0.030
β_8	0.0	0.001	0.003	-0.044	0.043	0.029	0.029
σ	1.0	1.013	1.009	0.927	1.106	0.060	0.061
φ_0	-1.0	-1.089	-1.085	-1.348	-0.861	0.150	0.174
φ_1	0.3	0.352	0.348	0.149	0.515	0.118	0.129
φ_{21}	0.3	0.289	0.323	-0.483	1.031	0.469	0.467
φ_{22}	0.3	0.348	0.338	-0.212	0.789	0.281	0.283
φ_{23}	0.3	0.369	0.362	-0.062	0.735	0.246	0.254
φ_{24}	0.3	0.283	0.256	-0.176	0.731	0.284	0.283
φ_{25}	0.3	0.353	0.343	-0.085	0.769	0.260	0.264
φ_{26}	0.3	0.358	0.326	-0.129	0.770	0.262	0.267
φ_{27}	0.3	0.364	0.352	-0.147	0.759	0.280	0.286
φ_{28}	0.3	0.292	0.270	-0.117	0.739	0.261	0.260

TABLE 4: The variable selection proportions results for β in the first simulation study.

Case	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
(i) Empirical Bayesian scenario								
i	0.00	0.00	0.96	0.97	0.00	0.95	0.94	0.98
ii	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
iii	0.00	0.96	0.98	0.93	0.94	0.96	0.97	0.94
(ii) Hierarchical model scenario								
i	0.00	0.00	0.99	0.98	0.00	1.00	1.00	1.00
ii	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
iii	0.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00

$$\text{logit}\{\text{pr}(r_i = 1|y_i, \mathbf{x}_i, \boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 y_i + \varphi_{21} x_{i1} + \varphi_{22} x_{i2} + \varphi_{23} x_{i3}, \quad (21)$$

with $\varphi_0 = -1.0$ and $\varphi_1 = \varphi_{21} = \varphi_{22} = \varphi_{23} = 0.3$.

Type ii: missing at random (MAR) missingness mechanism.

Type iii: nonignorable missingness mechanism model specified in M_0 .

TABLE 5: Bayesian estimates for case i in the second simulation study.

Parameter	True	Mean	Median	5%	95%	SD	RMSE
(i) Empirical Bayesian scenario							
β_1	3.0	2.988	2.997	2.846	3.096	0.083	0.083
β_2	1.5	1.482	1.473	1.328	1.638	0.089	0.091
β_3	0.0	-0.005	-0.001	-0.098	0.118	0.051	0.051
β_4	0.0	0.006	0.000	-0.108	0.095	0.063	0.063
β_5	2.0	1.992	1.992	1.858	2.124	0.092	0.092
β_6	0.0	-0.008	-0.001	-0.158	0.056	0.064	0.064
β_7	0.0	-0.001	-0.000	-0.095	0.079	0.056	0.055
β_8	0.0	0.002	0.000	-0.098	0.118	0.051	0.051
σ	1.0	0.989	0.986	0.914	1.079	0.058	0.058
φ_0	-2.5	-2.564	-2.567	-2.689	-2.452	0.072	0.096
φ_1	0.3	0.339	0.345	0.104	0.525	0.135	0.140
φ_{21}	0.3	0.282	0.261	-0.395	0.850	0.413	0.411
φ_{22}	0.3	0.274	0.258	-0.261	0.801	0.355	0.354
φ_{23}	0.3	0.300	0.287	-0.205	0.710	0.284	0.282
φ_{24}	0.3	0.331	0.329	-0.068	0.731	0.266	0.267
φ_{25}	0.3	0.283	0.279	-0.269	0.762	0.327	0.326
φ_{26}	0.3	0.293	0.294	-0.255	0.698	0.304	0.302
φ_{27}	0.3	0.291	0.311	-0.196	0.739	0.282	0.281
φ_{28}	0.3	0.387	0.337	-0.015	0.853	0.286	0.298
(ii) Hierarchical model scenario							
β_1	3.0	2.984	2.983	2.822	3.129	0.101	0.102
β_2	1.5	1.506	1.510	1.345	1.654	0.101	0.101
β_3	0.0	0.008	0.007	-0.067	0.089	0.047	0.047
β_4	0.0	-0.009	-0.006	-0.082	0.041	0.043	0.043
β_5	2.0	1.991	1.976	1.848	2.139	0.101	0.101
β_6	0.0	-0.001	0.003	-0.080	0.048	0.042	0.042
β_7	0.0	0.001	-0.000	-0.075	0.077	0.046	0.046
β_8	0.0	0.007	0.004	-0.054	0.064	0.041	0.042
σ	1.0	0.998	0.995	0.921	1.086	0.055	0.054
φ_0	-2.5	-2.573	-2.576	-2.681	-2.470	0.070	0.101
φ_1	0.3	0.327	0.315	0.133	0.501	0.119	0.121
φ_{21}	0.3	0.287	0.276	-0.340	0.756	0.334	0.333
φ_{22}	0.3	0.268	0.282	-0.260	0.710	0.320	0.320
φ_{23}	0.3	0.342	0.314	-0.126	0.834	0.293	0.294
φ_{24}	0.3	0.307	0.321	-0.207	0.797	0.308	0.307
φ_{25}	0.3	0.250	0.221	-0.251	0.735	0.281	0.284
φ_{26}	0.3	0.340	0.354	-0.125	0.712	0.283	0.285
φ_{27}	0.3	0.323	0.286	-0.187	0.847	0.336	0.335
φ_{28}	0.3	0.297	0.311	-0.185	0.707	0.281	0.279

FIGURE 1: Box plots summarizing the MAD for cases under (a) empirical Bayesian scenario (left panel) and (b) hierarchical model scenario (right panel) in the first and second simulation studies.

TABLE 6: Bayesian estimates for type i in the example 3.

Parameter	True	Mean	Median	5%	95%	SD	RMSE
(i) Empirical Bayesian scenario							
β_1	3.0	3.005	2.984	2.842	3.192	0.111	0.111
β_2	1.5	1.502	1.509	1.305	1.639	0.111	0.111
β_3	0.0	0.004	0.000	-0.100	0.120	0.077	0.077
β_4	0.0	-0.004	-0.001	-0.107	0.072	0.061	0.061
β_5	2.0	1.978	1.988	1.815	2.129	0.101	0.103
β_6	0.0	0.014	0.001	-0.039	0.091	0.054	0.056
β_7	0.0	0.002	0.000	-0.100	0.112	0.057	0.057
β_8	0.0	0.005	0.001	-0.062	0.097	0.058	0.058
σ	1.0	0.994	0.989	0.887	1.081	0.059	0.059
φ_0	-1.0	-1.087	-1.089	-1.383	-0.819	0.174	0.193
φ_1	0.3	0.366	0.384	0.087	0.593	0.159	0.171
φ_{21}	0.3	0.336	0.324	-0.410	0.955	0.441	0.440
φ_{22}	0.3	0.258	0.292	-0.448	0.759	0.372	0.371
φ_{23}	0.3	0.357	0.329	0.038	0.667	0.211	0.217
φ_{24}	0.0	0.027	-0.007	-0.380	0.437	0.260	0.260
φ_{25}	0.0	-0.066	-0.098	-0.677	0.478	0.337	0.342
φ_{26}	0.0	0.013	-0.015	-0.456	0.311	0.249	0.248
φ_{27}	0.0	-0.018	-0.056	-0.487	0.410	0.282	0.281
φ_{28}	0.0	-0.015	-0.004	-0.393	0.348	0.247	0.246
(ii) Hierarchical model scenario							
β_1	3.0	3.010	3.001	2.813	3.185	0.120	0.120
β_2	1.5	1.468	1.457	1.282	1.633	0.110	0.114
β_3	0.0	0.028	0.018	-0.148	0.165	0.103	0.106
β_4	0.0	-0.004	0.002	-0.191	0.150	0.109	0.108
β_5	2.0	1.991	1.969	1.812	2.218	0.126	0.126
β_6	0.0	-0.003	-0.001	-0.178	0.142	0.102	0.102
β_7	0.0	-0.007	-0.008	-0.161	0.140	0.104	0.103
β_8	0.0	0.002	-0.007	-0.157	0.165	0.094	0.093
σ	1.0	1.019	1.020	0.908	1.108	0.063	0.066
φ_0	-1.0	-1.052	-1.055	-1.290	-0.849	0.143	0.151
φ_1	0.3	0.319	0.300	0.031	0.581	0.175	0.175
φ_{21}	0.3	0.379	0.351	-0.494	1.081	0.503	0.507
φ_{22}	0.3	0.371	0.364	-0.229	0.907	0.338	0.344
φ_{23}	0.3	0.343	0.314	-0.082	0.709	0.237	0.240
φ_{24}	0.0	0.021	0.012	-0.381	0.426	0.255	0.255
φ_{25}	0.0	0.020	0.006	-0.601	0.566	0.372	0.371
φ_{26}	0.0	0.032	0.038	-0.497	0.399	0.272	0.271
φ_{27}	0.0	-0.054	-0.071	-0.512	0.351	0.270	0.275
φ_{28}	0.0	0.074	0.065	-0.285	0.413	0.217	0.229

For missing data simulated via type i and type ii, we obtain the Bayesian estimates under incorrect logistic regression model M_0 . For missing data simulated via type iii, we calculate the Bayesian estimates under incorrect MAR missingness mechanism. Summaries of these results are reported in Tables 6–9. Inspecting these results, the following observations are indicated: (1) results produced using the model M_0 are quite robust and accurate even if the true missing data are MAR, or the true missingness mechanisms are much simpler models; (2) performances obtained on the basis of the incorrect MAR assumption are poor. These observations agreed with those obtained by Lee and Tang [24].

4.4. A Real Example. In order to further demonstrate our proposed Bayesian methodologies, we consider a real example from the air pollution data in this subsection, which is available at the website <http://lib.stat.cmu.edu/datasets/>

NO2.data. The air pollution data includes $n = 500$ randomly chosen responses and $p = 7$ explanatory variables. For $i = 1, 2, \dots, n$, the response variable y_i is hourly values of the logarithm of the concentration of NO2, measured at Alnabru in Oslo, Norway, between October 2001 and August 2003; the explanatory variable x_{i1} denotes the logarithm of the number of cars per hour, x_{i2} represents temperature two meters above ground, x_{i3} is wind speed, x_{i4} denotes the temperature difference between 25 and 2 m above ground, x_{i5} represents wind direction, and x_{i6} and x_{i7} are hours of day and day number from October 1st, 2001, respectively. Each explanatory variable is standardized to have mean 0 and variance 1 during exploratory data analysis. The air pollution data has been analyzed and fitted by Zhu and Zhu [26], and they observed that all explanatory variables except x_{i5} and x_{i7} make a significant contribution to y_i . Model equation (19) and the following model

TABLE 7: Bayesian estimates for type ii in the example 3.

Parameter	True	Mean	Median	5%	95%	SD	RMSE
(i) Empirical Bayesian scenario							
β_1	3.0	2.970	2.976	2.745	3.118	0.110	0.113
β_2	1.5	1.488	1.477	1.324	1.639	0.109	0.110
β_3	0.0	-0.003	-0.001	-0.129	0.084	0.067	0.066
β_4	0.0	0.008	-0.000	-0.039	0.086	0.056	0.057
β_5	2.0	1.968	1.974	1.793	2.100	0.094	0.099
β_6	0.0	0.004	0.000	-0.083	0.074	0.110	0.113
β_7	0.0	0.003	-0.000	-0.081	0.087	0.050	0.050
β_8	0.0	0.005	0.000	-0.084	0.095	0.049	0.049
σ	1.0	1.001	0.996	0.894	1.106	0.064	0.064
φ_0	—	-0.622	-0.613	-0.888	-0.378	0.160	—
φ_1	—	-0.052	-0.057	-0.317	0.168	0.151	—
φ_{21}	—	0.199	0.194	-0.602	0.894	0.476	—
φ_{22}	—	0.109	0.103	-0.412	0.598	0.314	—
φ_{23}	—	-0.013	-0.022	-0.383	0.311	0.223	—
φ_{24}	—	-0.012	-0.001	-0.390	0.304	0.217	—
φ_{25}	—	0.107	0.098	-0.493	0.646	0.335	—
φ_{26}	—	0.015	0.026	-0.333	0.265	0.120	—
φ_{27}	—	0.006	-0.010	-0.381	0.337	0.215	—
φ_{28}	—	0.024	0.013	-0.286	0.364	0.204	—
(ii) Hierarchical model scenario							
β_1	3.0	2.974	2.969	2.816	3.107	0.095	0.098
β_2	1.5	1.479	1.486	1.261	1.656	0.123	0.124
β_3	0.0	0.004	0.000	-0.066	0.070	0.047	0.047
β_4	0.0	0.007	0.001	-0.066	0.099	0.056	0.056
β_5	2.0	1.973	1.968	1.806	2.118	0.101	0.104
β_6	0.0	0.008	0.003	-0.068	0.098	0.054	0.055
β_7	0.0	0.009	0.004	-0.051	0.047	0.049	0.050
β_8	0.0	-0.003	-0.003	-0.069	0.040	0.049	0.049
σ	1.0	1.014	1.011	0.903	1.115	0.066	0.068
φ_0	—	-0.606	-0.603	-0.842	-0.374	0.152	—
φ_1	—	-0.054	-0.078	-0.361	0.168	0.152	—
φ_{21}	—	0.168	0.135	-0.552	0.863	0.459	—
φ_{22}	—	0.097	0.075	-0.429	0.601	0.306	—
φ_{23}	—	0.012	0.037	-0.342	0.296	0.207	—
φ_{24}	—	-0.006	-0.029	-0.316	0.232	0.176	—
φ_{25}	—	0.135	0.157	-0.522	0.697	0.344	—
φ_{26}	—	-0.037	-0.049	-0.375	0.299	0.200	—
φ_{27}	—	0.050	0.050	-0.348	0.393	0.234	—
φ_{28}	—	-0.033	0.029	-0.397	0.267	0.202	—

$$\text{logit}\{\text{pr}(r_i = 1|y_i, \mathbf{x}_i, \boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 y_i + \varphi_{21} x_{i1} + \cdots + \varphi_{27} x_{i7}, \quad (22)$$

where $\varphi_0 = -1.0$, $\varphi_1 = \varphi_{21} = \cdots = \varphi_{27} = 0.5$ are used to reanalyze the air pollution data in the article, where the logistic regression model (22) may represent that we create artificial missing data by deliberated deleting some of responses according to the given missing mechanism, and the missing data's proportion generated in this way is 0.326. Then, we employ the preceding propose Bayesian procedure to make statistical inferences on the assumption of the above model settings. Here, 10 000 samples after 10 000 burn-in iterations are employed to calculate the Bayesian estimates of unknown parameters. In M-H algorithm, we take $\sigma_y^2 = 2.2$ and $\sigma_\varphi^2 = 1.0$ giving average acceptance rates 29.0% and 26.8%, respectively. Also, we conduct a real data analysis

under the assumption for the fully observed data in the framework of Bayesian adaptive Lasso.

Table 9 reports the corresponding results for regression coefficients by two different methods, in which 'EB' and 'HM' denote the Bayesian estimates on the basis of Empirical Bayesian method and Hierarchical model method, respectively. From Table 9, we clearly see that (1) the estimate results for the regression coefficients in all two methods are quite close; (2) almost all standard errors obtained under the fully observed data are smaller than those obtained under the missing data, which indicates that the procedures under the fully observed data outperform the procedures under the missing data, but the differences for performances are minor; (3) all these procedures could identify that non-zero significant regression coefficients are $\beta_1, \beta_2, \beta_3, \beta_4$, and β_6 , which is consistent with the results given in Zhu and Zhu

TABLE 8: Bayesian estimates for type iii in the example 3.

Parameter	True	Mean	Median	5%	95%	SD	RMSE
(i) Empirical Bayesian scenario							
β_1	3.0	2.900	2.982	0.001	3.606	0.737	0.740
β_2	1.5	1.320	1.437	0.000	1.799	0.644	0.666
β_3	0.0	0.103	0.003	-0.085	0.617	0.287	0.304
β_4	0.0	0.055	0.000	-0.082	0.181	0.240	0.245
β_5	2.0	1.745	1.964	0.001	2.128	0.640	0.686
β_6	0.0	0.061	0.000	-0.067	0.464	0.210	0.218
β_7	0.0	0.007	-0.000	-0.013	0.019	0.042	0.043
β_8	0.0	-0.003	-0.000	-0.052	0.021	0.035	0.035
σ	1.0	1.270	1.033	0.889	2.305	0.532	0.594
(ii) Hierarchical model scenario							
β_1	3.0	3.195	3.185	2.974	3.395	0.140	0.240
β_2	1.5	1.629	1.606	1.414	1.881	0.143	0.192
β_3	0.0	0.050	0.034	-0.127	0.300	0.131	0.139
β_4	0.0	0.068	0.048	-0.164	0.299	0.126	0.142
β_5	2.0	2.139	2.135	1.900	2.354	0.137	0.195
β_6	0.0	0.053	0.040	-0.157	0.270	0.123	0.133
β_7	0.0	0.052	0.048	-0.147	0.245	0.128	0.137
β_8	0.0	0.063	0.058	-0.097	0.217	0.099	0.117
σ	1.0	1.250	1.242	1.145	1.364	0.075	0.261

TABLE 9: Bayesian estimates and standard errors (in the parentheses) of parameter β in the air pollution data analysis.

	Missing data		Fully observed data	
	EB	HM	EB	HM
β_1	0.547 (0.038)	0.552 (0.039)	0.541 (0.031)	0.540 (0.031)
β_2	-0.156 (0.035)	-0.163 (0.036)	-0.146 (0.028)	-0.145 (0.029)
β_3	-0.273 (0.035)	-0.268 (0.036)	-0.225 (0.025)	-0.225 (0.025)
β_4	0.179 (0.038)	0.200 (0.041)	0.172 (0.028)	0.171 (0.028)
β_5	0.034 (0.028)	0.046 (0.035)	0.034 (0.025)	0.033 (0.028)
β_6	-0.121 (0.038)	-0.121 (0.040)	-0.123 (0.030)	-0.122 (0.031)
β_7	0.000 (0.002)	0.053 (0.035)	0.039 (0.025)	0.037 (0.027)

[26]. In the end, PPP p -value for goodness-of-fit with respect to model (19) with the nonignorable missing mechanism defined in model (20) is 0.446 and 0.454 based on Empirical Bayesian method and Hierarchical model method, respectively, which illustrates that our considered models as well as the proposed goodness-of-fit statistic are feasible.

5. Conclusion and Discussion

We develop a Bayesian adaptive lasso approach to analyze linear regression models with nonignorable missing responses in this article, where the nonignorable missingness mechanism is specified by a logistic regression model. With the help of Gibbs sampler [14] and the Metropolis-Hastings algorithm [15, 16], we deal with high dimensional integral with not having a closed form in the posterior analysis that is often computationally intractable and then obtain the Bayesian estimates of unknown parameters and their corresponding standard error estimates, the PPP p value [23] for goodness-of-fit. In short, empirical performances of proposed Bayesian approaches in our simulation studies are feasible and effective, and applications for the newly developed statistical tools to the proposed Bayesian

adaptive lasso with nonignorable missing responses are plausible.

Data Availability

The data used to support the findings of simulation studies are available from the corresponding author upon request. The air pollution data in the real example are from previously reported studies and datasets, which are available at the website <http://lib.stat.cmu.edu/datasets/NO2.data>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (11761016; 12161014), Science and Technology Foundation of Guizhou Province of China ([2020]1Y009), Project of High Level Creative Talents in Guizhou Province of China, the Natural Science Foundation of the Education Department of Guizhou province in China ([2020]090), and Discipline and Master's Site Construction

Project of Guiyang University by Guiyang City Financial Support Guiyang University (SX-2020).

References

- [1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [2] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression (with discussion)," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–451, 2004.
- [3] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [4] M. Yun and Y. Lin, "Model selection and estimation in regression with grouped variables," *JR Stat Soc Ser B*, vol. 68, no. 1, pp. 49–67, 2006.
- [5] N. Meinshausen and P. Bühlmann, "Stability selection," *Journal of the Royal Statistical Society: Series B*, vol. 72, no. 4, pp. 417–473, 2010.
- [6] T. Park and G. Casella, "The bayesian lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.
- [7] R. Guo, H. Zhu, S.-M. Chow, and J. G. Ibrahim, "Bayesian lasso for semiparametric structural equation models," *Biometrics*, vol. 68, no. 2, pp. 567–577, 2012.
- [8] C. Hans, "Model uncertainty and variable selection in Bayesian lasso regression," *Statistics and Computing*, vol. 20, no. 2, pp. 221–229, 2010.
- [9] C. Leng, M.-N. Tran, and D. Nott, "Bayesian adaptive lasso," *Annals of the Institute of Statistical Mathematics*, vol. 66, no. 2, pp. 221–244, 2014.
- [10] R. J. A. Little and D. B. Rubin, *Statistical Analysis With missing Data*, Wiley, New York, NY, USA, 2nd edition, 2002.
- [11] J. G. Ibrahim, H. Zhu, and N. Tang, "Model selection criteria for missing-data problems using the EM algorithm," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1648–1658, 2008.
- [12] R. I. Garcia, J. G. Ibrahim, and H. Zhu, "Variable selection for Regression models With Missing data," *Statistica Sinica*, vol. 20, no. 1, pp. 149–165, 2010.
- [13] P. X. Zhao and L. G. Xue, "Variable selection for semi-parametric varying-coefficient partially linear models with missing response at random," *Acta Mathematica Sinica, English Series*, vol. 27, no. 11, pp. 2205–2216, 2011.
- [14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984.
- [15] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machine," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1091, 1953.
- [17] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation (with discussion)," *J Amer Statist Assoc*, vol. 82, no. 398, pp. 528–540, 1987.
- [18] A. Gelman, "Inference and monitoring convergence," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds., Chapman and Hall, London, UK, 1996.
- [19] Y. Y. Zhao and N. S. Tang, "Maximum-likelihood estimation and influence analysis in multivariate skew-normal reproductive dispersion mixed models for longitudinal data," *Statistics*, vol. 49, no. 6, pp. 1348–1365, 2015.
- [20] D. Xu, Z. Zhang, and L. Wu, "Bayesian analysis for semi-parametric mixed -effects double regression models," *Hacet J Math Stat*, vol. 45, no. 1, pp. 279–296, 2016.
- [21] A. Gelman, G. O. Roberts, and W. R. Gilks, "Efficient metropolis jumping rules," in *Bayesian Statistics*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds., Oxford University Press, Oxford, UK, 1995.
- [22] C. J. Geyer, "Practical Markov chain Monte Carlo," *Statistical Science*, vol. 7, no. 4, pp. 473–483, 1992.
- [23] M. J. Bayarri and J. O. Berger, "PValues for composite null models," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1127–1142, 2000.
- [24] S.-Y. Lee and N. S. Tang, "Bayesian analysis of nonlinear structural equation models with nonignorable missing data," *Psychometrika*, vol. 71, no. 3, pp. 541–564, 2006.
- [25] N. S. Tang and Y. Y. Zhao, "Semiparametric Bayesian analysis of nonlinear reproductive dispersion mixed models for longitudinal data," *Journal of Multivariate Analysis*, vol. 115, no. 1, pp. 68–83, 2013.
- [26] L.-P. Zhu and L.-X. Zhu, "Nonconcave penalized inverse regression in single-index models with high dimensional predictors," *Journal of Multivariate Analysis*, vol. 100, no. 5, pp. 862–875, 2009.