



## A Bayesian hierarchical model for quantitative and qualitative responses

Lulu Kang, Xiaoning Kang, Xinwei Deng & Ran Jin

To cite this article: Lulu Kang, Xiaoning Kang, Xinwei Deng & Ran Jin (2018) A Bayesian hierarchical model for quantitative and qualitative responses, Journal of Quality Technology, 50:3, 290-308, DOI: [10.1080/00224065.2018.1489042](https://doi.org/10.1080/00224065.2018.1489042)

To link to this article: <https://doi.org/10.1080/00224065.2018.1489042>



Published online: 18 Sep 2018.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)

RESEARCH PAPER



## A Bayesian hierarchical model for quantitative and qualitative responses

Lulu Kang<sup>a</sup>, Xiaoning Kang<sup>b</sup>, Xinwei Deng<sup>c</sup>, and Ran Jin<sup>d</sup>

<sup>a</sup>Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois; <sup>b</sup>Institute of Supply Chain Analytics and International Business College, Dongbei University of Finance and Economics, Dalian, China; <sup>c</sup>Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia; <sup>d</sup>Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

### ABSTRACT

In many science and engineering systems both quantitative and qualitative output observations are collected. If modeled separately the important relationship between the two types of responses is ignored. In this article, we propose a Bayesian hierarchical modeling framework to jointly model a continuous and a binary response. Compared with the existing methods, the Bayesian method overcomes two restrictions. First, it solves the problem in which the model size (specifically, the number of parameters to be estimated) exceeds the number of observations for the continuous response. We use one example to show how such a problem can easily occur if the design of the experiment is not proper; all the frequentist approaches would fail in this case. Second, the Bayesian model can provide statistical inference on the estimated parameters and predictions, whereas it is not clear how to obtain inference using the latest method proposed by Deng and Jin (2015), which jointly models the two responses via constrained likelihood. We also develop a Gibbs sampling scheme to generate accurate estimation and prediction for the Bayesian hierarchical model. Both the simulation and the real case study are shown to illustrate the proposed method.

### KEYWORDS

Bayesian hierarchical model; conjugate prior; generalized linear model; Gibbs sampling; qualitative response; quantitative response

## 1. Introduction

Many data collected from engineering and scientific systems contain both quantitative and qualitative (QQ) output observations or responses. For example, in the lapping stage of the wafer manufacturing process the qualitative response is the conformity of the site total indicator reading (STIR) of the wafer, which has two possible outcomes: whether or not the wafer STIR is within the tolerance. The quantitative response is the total thickness variation (TTV) of the wafer. Both of the response variables measure the smoothness of the wafers, which is an important geometrical quality index of the wafers. See Ning et al. (2012), Zhao et al. (2011), and Zhang et al. (2015) for detailed studies of these two quality characteristics. An interpretable and accurate statistical modeling approach is needed to find out how the controllable process variables and the covariates affect the two kinds of responses. Among all possible modeling techniques, the simplest approach is to model the two types of responses separately. We can use linear regression models for the quantitative response and

generalized linear models or classification methods for the qualitative response. But doing so would ignore the possible association between the two responses. Deng and Jin (2015) have shown the necessity for jointly modeling the two kinds of responses for the lapping process experiment. Such association is important as it provides us with some insightful understandings of the system under study as well as a significant improvement in the prediction accuracy.

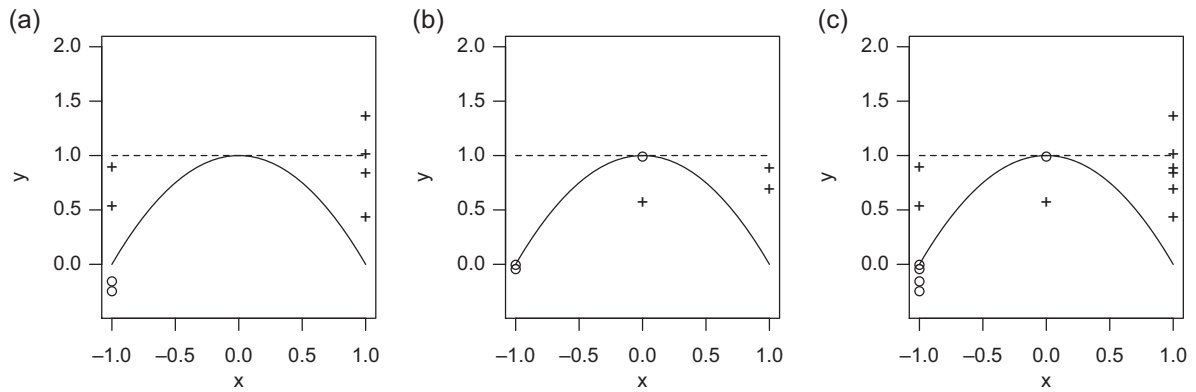
There is some existing literature on how to jointly model the mixed-type responses. One group of the existing methods directly use latent variables to model the correlations between different responses. Among them, Dunson and Herring (2005), Weiss et al. (2011), and Bello et al. (2012) are based on Bayesian framework. The others use the frequentist approach, including Catalano and Ryan (1992) and Sammel et al. (1997). Another group of methods factorize the joint distribution of the mixed-type responses, including Cox and Wermuth (1992), Fitzmaurice and Laird (1995, 1997), Chen et al. (2014), Yang et al. (2014), Deng and Jin (2015), and Guglielmi et al. (2016). Cox

**CONTACT** Lulu Kang ✉ [lkang2@iit.edu](mailto:lkang2@iit.edu) 📧 Department of Applied Mathematics, Illinois Institute of Technology, 10 West 32nd Street, Chicago, IL 60616, USA.

Lulu Kang and Xiaoning Kang share the first authorship.

Color versions of one or more figures in this article can be found online at [www.tandfonline.com/ujqt](http://www.tandfonline.com/ujqt).

© 2018 American Society for Quality



**Figure 1.** (a) Observations from design for  $Z$ ; (b) observations from design for  $Y$ ; (c) observations from the combined design. A dashed line “- -” denotes  $E(Y|Z=1) = 1$ ; a solid line “—” denotes  $E(Y|Z=0) = 1 - x^2$ ; point “+” denotes  $(x_i, y_i)$  with  $z_i = 1$ ; point “o” denotes  $(x_i, y_i)$  with  $z_i = 0$ . (See also Figure 1 in Kang et al. 2015.)

and Wermuth (1992) discussed two factorization models for a continuous and a binary response as the function of covariates. Fitzmaurice and Laird (1995, 1997) considered the marginal distribution of the qualitative response as well as the conditional distribution of the quantitative response and proposed a likelihood-based method. Chen et al. (2014) and Yang et al. (2014) introduced the mixed graphical model to analyze the association between the quantitative and qualitative responses. Guglielmi et al. (2016) proposed a semiparametric Bayesian joint model for one continuous and two binary responses. See Deng and Jin (2015) and Guglielmi et al. (2016) for more detailed reviews.

Based on the same factorization of the joint-likelihood idea, Deng and Jin (2015) assumed a conditional model for the quantitative response given the values of the qualitative response. Specifically, the quantitative response follows different normal distributions for different qualitative response values. To achieve model sparsity the authors used the nonnegative garrote and constrained likelihood estimation method. The authors also demonstrated that this approach provides parsimonious, interpretable, and accurate models for both responses. Despite these advantages, it has two main restrictions.

The first restriction is common to all frequentist approaches. We demonstrate it using a toy example. Denote by  $Y$  and  $Z$  a continuous response and a binary qualitative response, respectively. Assume that the true model of  $Z$  follows the Bernoulli distribution with  $E(Z|x) = \pi(x) = \exp(1+x)/(1+\exp(1+x))$  and the true model of  $Y$  is  $Y|Z = z \sim N(1 - (1-z)x^2, 0.3^2)$  (i.e.,  $E(Y|Z=1) = 1$  and  $E(Y|Z=0) = 1 - x^2$ ). We construct a 14-point design, which consists of an 8-point local  $D$ -optimal design for the logistic model  $\log(\pi(x)/(1-\pi(x))) = \eta_0 + \eta_1 x$  of  $Z$  and a 6-point  $D$ -optimal design for the quadratic regression model of  $Y$ . Figure 1(a)–(c) shows the data from the 8-point, 6-point, and their combined 14-point designs, respectively.

Clearly, if using the frequentist approach it is impossible to estimate the quadratic model of  $E(Y|Z=0)$  as there are no observations of  $Y|Z=0$  at  $x=1$ . But such a defect from the design of experiment can be solved by the Bayesian approach with an informative prior.

The second restriction is that statistical inference is hard to obtain from the joint model developed in Deng and Jin (2015). This joint QQ model is also introduced in Section 2.1. A constraint-likelihood approach was used to estimate parameters for the joint QQ models. The authors chose the nonnegative garrote approach (Breiman 1995) to enable variable selection for a parsimonious model with meaningful interpretation. More importantly, they incorporated the heredity principles (Wu and Hamada 2011) into the nonnegative garrote in the forms of constraints. Thus the estimation becomes the constrained maximum likelihood estimation (MLE). Such a complicated constrained likelihood optimization often does not have an analytical solution as in linear regression models. To solve this constrained optimization Deng and Jin (2015) developed the constrained sequential quadratic programming that iteratively searches for the optimal solution. Consequently, the classic asymptotical distribution of the maximum likelihood estimation cannot be easily applied here as in the ordinary generalized linear model. Thanks to the Bayesian framework, the inferences of the estimated parameters and the predictions are naturally available from the posterior distribution and the posterior predictive distribution. What is more, we also incorporate the hierarchical ordering principle (Wu and Hamada 2011) in the marginal prior components of the linear coefficients so as to achieve sparsity and meaningful interpretation.

The two restrictions on the existing frequentist methods motivate us to investigate the Bayesian framework as an alternative. Different from some existing Bayesian approaches, we do not consider using the

latent variable to model the correlation between the two responses because we prefer the clear model structure in Deng and Jin (2015). It is also different from the Bayesian method in Guglielmi et al. (2016), where the marginal distribution was applied to the continuous response, whereas our approach applies the marginal distribution to the binary response. We also propose a Bayesian hierarchical structure to link together all the conditional models for the continuous response and the marginal model for the binary response. The structure also has satisfactory performances in prediction and variable selection.

The rest of the article is outlined as follows. Section 2 introduces the Bayesian hierarchical modeling framework. In Section 3 a Gibbs sampling algorithm is developed for estimation and prediction. We use simulation studies to demonstrate the performance of the proposed method under different circumstances in Section 4; Section 5 considers the case study on the lapping process in wafer manufacturing and compares the results with previous results. The article concludes with Section 6. Our algorithm and data analysis are all implemented in R.

## 2. The Bayesian hierarchical model

### 2.1. Sampling distribution

Suppose that  $Y$  is a continuous response and  $Z$  is a binary response. The independent variable  $\mathbf{x} = (x_1, \dots, x_p)' \in \mathbb{R}^p$  has  $p$  dimensions. It can contain controllable process variables and uncontrollable covariates. To jointly model the two responses  $Y$  and  $Z$  given  $\mathbf{x}$ , consider the joint probability of  $Y|Z$  and  $Z$ . The conditional model of  $Y|Z$  is assumed to be a linear regression model, while the model of  $Z$  is a logistic regression model. Specifically, we consider a joint modeling of  $Y$  and  $Z$  as follows:

$$Z = \begin{cases} 1, & \text{with probability } \pi(\mathbf{x}) \\ 0, & \text{with probability } 1 - \pi(\mathbf{x}) \end{cases} \text{ with } \pi(\mathbf{x}, \boldsymbol{\eta}) \quad [1]$$

$$= \frac{\exp(\mathbf{f}(\mathbf{x})' \boldsymbol{\eta})}{1 + \exp(\mathbf{f}(\mathbf{x})' \boldsymbol{\eta})},$$

where  $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_q(\mathbf{x}))'$  contains  $q$  modeling effects including the intercept and main, interaction, and quadratic effects, etc., and  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_q)'$  is a vector of coefficient parameters. Conditional on  $Z = z$ , the quantitative variable  $Y$  has the distribution

$$Y|Z = z \sim N\left(z\mathbf{f}(\mathbf{x})' \boldsymbol{\beta}^{(1)} + (1 - z)\mathbf{f}(\mathbf{x})' \boldsymbol{\beta}^{(2)}, \sigma^2\right), \quad [2]$$

where  $\boldsymbol{\beta}^{(i)} = (\beta_1^{(i)}, \dots, \beta_q^{(i)})'$ ,  $i = 1, 2$  are the corresponding coefficients of the modeling effects. The parameter

$\sigma^2$  is the noise variance. The above QQ model indicates that  $Y|Z = 1 \sim N(\mathbf{f}(\mathbf{x})' \boldsymbol{\beta}^{(1)}, \sigma^2)$  and  $Y|Z = 0 \sim N(\mathbf{f}(\mathbf{x})' \boldsymbol{\beta}^{(2)}, \sigma^2)$ . The association between the two responses  $Y$  and  $Z$  is represented using the conditional model  $Y|Z$ . When the two linear models for  $Y|Z = 0$  and  $Y|Z = 1$  are different (i.e.,  $\boldsymbol{\beta}^{(1)} \neq \boldsymbol{\beta}^{(2)}$ ), then it is important to take into account the influence of the qualitative response  $Z$  when modeling the quantitative response  $Y$ . This conditional model structure of  $Y|Z$  and  $Z$ , called the joint QQ model, was also used in Deng and Jin (2015) and Kang et al. (2015).

Note that here we assume that the linear model for  $Y$  and the logistic model for  $Z$  involve the same modeling effects  $\mathbf{f}(\mathbf{x})$ . But if necessary we can assume different modeling effects for  $E(Y|Z = 1)$ ,  $E(Y|Z = 0)$ , and  $\text{logit}(\pi(\mathbf{x}))$ . In our model we let  $\mathbf{f}(\mathbf{x})$  contain all the necessary terms for both  $Y$  and  $Z$ . If some terms are statistically insignificant for a certain response, the posterior means of the corresponding coefficients should be shrunk and be close to zero, as in a Bayesian linear regression or ridge regression. Another simplification we make is to assume the same variance  $\sigma^2$  for both  $Y|Z = 1$  and  $Y|Z = 0$ . Users can easily extend the proposed model to different variances.

Denote the data as  $(\mathbf{x}_i, y_i, z_i)$ ,  $i = 1, \dots, n$  where  $y_i \in \mathbb{R}$  and  $z_i \in \{0, 1\}$ . The vectors  $\mathbf{y} = (y_1, \dots, y_n)'$  and  $\mathbf{z} = (z_1, \dots, z_n)'$  are the vectors of response observations. Suppose that there are  $n_1$  observed  $y_i$ 's when  $Z = 1$  and  $n_2$  observed  $y_i$ 's when  $Z = 0$ . Let  $n = n_1 + n_2$ . Based on Eqs. [1] and [2], we can express the sampling distributions as

$$\begin{aligned} \mathbf{y}|\mathbf{z}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \sigma^2 &\sim N(\mathbf{V}_1 \mathbf{F} \boldsymbol{\beta}^{(1)} + \mathbf{V}_2 \mathbf{F} \boldsymbol{\beta}^{(2)}, \sigma^2 \mathbf{I}_n), \\ \mathbf{z}|\boldsymbol{\eta} &\sim_{\text{ind}} \text{Bernoulli}(\pi(\mathbf{x}_i, \boldsymbol{\eta})) \text{ for } i = 1, \dots, n, \end{aligned} \quad [3]$$

where  $\mathbf{V}_1 = \text{diag}\{z_1, \dots, z_n\}$  is a diagonal matrix,  $\mathbf{I}_n$  is the  $n \times n$  identity matrix,  $\mathbf{V}_2 = \mathbf{I}_n - \mathbf{V}_1$ , and  $\mathbf{F}$  is the model matrix with the  $i$ th row as  $\mathbf{f}(\mathbf{x}_i)'$ . Let  $p(\cdot)$  denote a general density function. The sampling distribution of  $\mathbf{z}$  is

$$p(\mathbf{z}|\boldsymbol{\eta}) \propto \exp\left\{\sum_{i=1}^n [z_i \mathbf{f}(\mathbf{x}_i)' \boldsymbol{\eta} - \log(1 + \exp(\mathbf{f}(\mathbf{x}_i)' \boldsymbol{\eta}))]\right\}.$$

### 2.2. Prior and hyperprior distributions

The unknown parameters in the sampling distribution are  $\boldsymbol{\theta} = (\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta}, \sigma^2)$ . Due to the normal distribution of  $\mathbf{y}$  the semiconjugate marginal prior components of  $\boldsymbol{\beta}^{(1)}$  and  $\boldsymbol{\beta}^{(2)}$  should come from the normal distribution family. The semiconjugate marginal prior component for  $\boldsymbol{\eta}$  is not normal (Chen and Ibrahim 2003), but we still

prefer to use the nonconjugate normal distribution as the marginal prior for  $\boldsymbol{\eta}$ . In this way, all the coefficients of the model terms  $f(\mathbf{x})$  have marginal prior components of the same format, thus making it simpler to specify the hyperprior distribution of the hyperparameters. In the context of this article we consider  $\sigma^2$  a nuisance parameter, so we choose a weakly informative marginal prior,  $\text{Inv} - \chi^2(0.001, 0.001)$ , which is a commonly considered vague proper prior distribution for variance parameters (Chapter 5, Gelman et al. 2014). To sum up, the marginal prior components for  $\boldsymbol{\beta}^{(1)}$ ,  $\boldsymbol{\beta}^{(2)}$ ,  $\boldsymbol{\eta}$ , and  $\sigma^2$  are

$$\begin{aligned} \boldsymbol{\beta}^{(i)} &\sim N(\mathbf{0}, \tau_i^2 \mathbf{R}_i) \text{ for } i = 1, 2, \\ \boldsymbol{\eta} &\sim N(\mathbf{0}, \tau_2^2 \mathbf{R}_3), \\ \sigma^2 &\sim \text{Inv} - \chi^2(0.001, 0.001). \end{aligned} \tag{4}$$

The matrices  $\mathbf{R}_1$ ,  $\mathbf{R}_2$ , and  $\mathbf{R}_3$  are the correlation matrices for the normal distributions and are explained later. The marginal prior components for  $\boldsymbol{\beta}^{(1)}$  and  $\boldsymbol{\beta}^{(2)}$  are independent for simplicity and are also based on the assumption that two,  $Y|Z=1$  and  $Y|Z=0$ , are not correlated. Under certain circumstances it might be necessary to assume that  $\boldsymbol{\beta}^{(1)}$  and  $\boldsymbol{\beta}^{(2)}$  are correlated in the prior. The proposed Bayesian hierarchical model can still be applied, but the full-conditional distribution of  $\boldsymbol{\beta}^{(1)}$  and  $\boldsymbol{\beta}^{(2)}$ —given the rest of the parameters and data—has a slightly more complicated covariance matrix. The prior mean  $\boldsymbol{\eta}$  is such that

$\Pr(Z = 1 | \mathbf{x}, E(\boldsymbol{\eta})) = 1/2$ . To avoid too many hyperparameters we also assume the prior mean of  $\boldsymbol{\beta}^{(1)}$  and  $\boldsymbol{\beta}^{(2)}$  to be zero. It is reasonable as long as we center  $\mathbf{y}$  in the data preprocessing stage. Let  $\bar{y}_1$  be the sample mean of  $y_i$ 's for  $z_i=1$  and  $\bar{y}_2$  for  $z_i=0$ . We center the  $Y$  observations as  $y_i - \bar{y}_1$  if  $z_i=1$  and  $y_i - \bar{y}_2$  if  $z_i=0$ , and we use the same notation  $\mathbf{y}$  for the centralized observations. Again for simplicity, we assume the same variance  $\tau_1^2$  in the marginal prior components of  $\boldsymbol{\beta}^{(1)}$  and  $\boldsymbol{\beta}^{(2)}$ . The variance can be easily changed to different values.

The correlation matrices in Eq. [4] in the marginal prior components for  $\boldsymbol{\beta}^{(1)}$ ,  $\boldsymbol{\beta}^{(2)}$ , and  $\boldsymbol{\eta}$  are assumed to be diagonal, which means that the coefficients are independent of each other. This assumption is reasonable if we use the orthogonal polynomial basis of  $\mathbf{x}$ , which consists of the intercept, the linear effects, the quadratic effects, and the interactions, etc., up to a user-specified order. If the controllable variable settings are from a full factorial design or an orthogonal design, then we can achieve full or near orthogonality between the bases. For the bases involving covariates, achieving full or near orthogonality is not likely. But we still assume independence for simplicity and leave the posterior distribution to correct it. Let  $\mathbf{R}_i = \text{diag}\{1, r_i, \dots, r_i, r_i^2, \dots, r_i^2, \dots\}$  for  $i = 1, 2, 3$  where  $r_i$

$\in (0, 1)$  is a hyperparameter. The power index of  $r_i$  is the same as the order of the corresponding polynomial term. For example, if the model  $f(\mathbf{x})$  for  $\mathbf{x} \in \mathbb{R}^2$  is a full quadratic model and contains the term  $\{1, x_1, x_2, x_1^2, x_2^2, x_1x_2\}$  then the corresponding prior correlation matrix should be specified as  $\mathbf{R} = \text{diag}\{1, r, r, r, r^2, r^2, r^2\}$ . In this way the prior variance of the effect decreases exponentially as the order of the effect increases, following the hierarchy ordering principle defined in Wu and Hamada (2011). The hierarchy ordering principle can reduce the size of the model and avoid including higher-order and less-significant model terms. Such prior distribution was first proposed by Joseph (2006) and later used by Kang and Joseph (2009) and Ai et al. (2009).

The hyperparameters in the marginal prior components of Eq. [4] are  $\boldsymbol{\phi} = (r_1, r_2, r_3, \tau_1^2, \tau_2^2)$ . We use the following marginal hyperprior components:

$$\begin{aligned} \tau_1^2, \tau_2^2 &\sim_{iid} \text{Inv} - \chi^2(\nu, \delta^2), \\ r_1, r_2, r_3 &\sim_{iid} \text{Beta}(a, b). \end{aligned}$$

Here  $\text{Inv} - \chi^2(\nu, \delta^2)$  denotes the scaled inverse Chi-square distribution with  $\nu$  degrees of freedom and scale  $\delta$ . For  $\tau_i^2$ , the inverse- $\chi^2$  distribution is semiconjugate with normal prior distributions of  $\boldsymbol{\beta}^{(1)}$ ,  $\boldsymbol{\beta}^{(2)}$ , and  $\boldsymbol{\eta}$  because the full-conditional distribution of  $\tau_i^2$  is also an inverse- $\chi^2$  distribution. As  $r_i \in (0, 1)$ , the beta distribution is an obvious choice. The parameters  $(\nu, \delta^2, a, b)$  can be considered as tuning parameters for the Bayesian hierarchical model.

### 3. Monte Carlo Markov Chain (MCMC) sampling

Given the model assumptions, the next step in a standard Bayesian approach is to sample the posterior distribution of the parameters and posterior predictive distribution of the response variables. The mode values of these distributions are usually considered as the parameter estimation or predictions at the query points, and statistical inference such as confidence intervals can be obtained from these empirical distributions.

Our goal is to sample from the joint posterior distribution  $p(\boldsymbol{\theta}, \boldsymbol{\phi} | \mathbf{y}, \mathbf{z})$ . For a relatively complicated Bayesian framework, the Gibbs sampling algorithm, a popular MCMC sampling method, can sequentially update each block of parameters following the corresponding full-conditional distribution. To briefly explain the Gibbs sampler let us use the general notation of parameter vector  $\boldsymbol{\theta}$  and data  $\mathbf{y}$ . The vector  $\boldsymbol{\theta}$  can be divided into  $s$  blocks (i.e.,  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_s)$ ). In the  $t$ th iteration of the MCMC chain each  $\boldsymbol{\theta}_{t,j}$  for  $j=1, \dots, s$  is sampled from the conditional distribution  $p(\boldsymbol{\theta}_j | \boldsymbol{\theta}_{t-1,-j}, \mathbf{y})$  where  $\boldsymbol{\theta}_{t-1,-j} = (\boldsymbol{\theta}_{t,1}, \dots, \boldsymbol{\theta}_{t,j-1}, \boldsymbol{\theta}_{t-1,j+1}, \dots, \boldsymbol{\theta}_{t-1,s})$ . For a more detailed



illustration of a Gibbs sampler and its convergence condition readers can turn to Casella and George (1992) and Roberts and Smith (1994). To develop the Gibbs sampler for the Bayesian hierarchical model we first derive the full-conditional distributions for the blocks of parameters in  $(\theta, \phi)$  in Section 3.1 and elaborate the Gibbs sampler in Section 3.2.

### 3.1. Full-conditional distributions

According to the sampling distribution, the prior, and hyperprior distribution, we can derive the full-conditional distributions for the parameters and the hyperparameters as follows. Many details on the derivation are omitted to save space. The derivation follows the classic Bayesian theory on deriving posterior distributions, which can be found in any Bayesian statistics textbook such as Gelman et al. (2014).

Due to the semiconjugacy it is straightforward to derive the full-conditional distribution for the linear coefficients  $\beta^{(1)}$  and  $\beta^{(2)}$  following the formula on Bayesian linear regression:

$$\begin{aligned} (\beta^{(1)}, \beta^{(2)}) | \eta, \sigma^2, \phi, \mathbf{y}, \mathbf{z} &\sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \text{ where} \\ \boldsymbol{\mu}_\beta &= \mathbf{V}_n \begin{bmatrix} \mathbf{F}' \mathbf{V}_1 \mathbf{y} \\ \mathbf{F}' \mathbf{V}_2 \mathbf{y} \end{bmatrix}, \\ \boldsymbol{\Sigma}_\beta = \sigma^2 \mathbf{V}_n &= \sigma^2 \begin{bmatrix} \mathbf{F}' \mathbf{V}_1 \mathbf{F} + \sigma^2 / \tau_1^2 \mathbf{R}_1^{-1}, & \mathbf{0} \\ \mathbf{0}, & \mathbf{F}' \mathbf{V}_2 \mathbf{F} + \sigma^2 / \tau_2^2 \mathbf{R}_2^{-1} \end{bmatrix}^{-1}. \end{aligned}$$

The full-conditional distribution of  $\sigma^2$  is an inverse- $\chi^2$  distribution according to the Bayesian linear regression model (Chapter 14 of Gelman et al. 2014). Thus,

$$\sigma^2 | \beta^{(1)}, \beta^{(2)}, \mathbf{y}, \mathbf{z} \sim \text{Inv} - \chi^2 \left( n + 0.001, \frac{\sum_{i=1}^n (y_i - \mu_i)^2 + 10^{-6}}{n + 0.001} \right),$$

where  $\mu_i$ 's are the elements of  $\boldsymbol{\mu} = \mathbf{V}_1 \mathbf{F} \beta^{(1)} + \mathbf{V}_2 \mathbf{F} \beta^{(2)}$ .

Because the normal marginal prior for  $\boldsymbol{\eta}$  is not semiconjugate with the sampling distribution of  $\mathbf{z}$ , the exact full-conditional distribution  $p(\boldsymbol{\eta} | \mathbf{z}, r_3, \tau_2^2)$  in Eq. [5] does not belong to any known distribution family:

$$\begin{aligned} p(\boldsymbol{\eta} | r_3, \tau_2^2, \mathbf{z}) &\propto (2\pi\tau_2^2)^{-p/2} \det(\mathbf{R}_3)^{-p/2} h(\boldsymbol{\eta} | \mathbf{z}, r_3, \tau_2^2), \\ &\propto h(\boldsymbol{\eta} | \mathbf{z}, r_3, \tau_2^2) \end{aligned} \quad [5]$$

where

$$\begin{aligned} h(\boldsymbol{\eta} | \mathbf{z}, r_3, \tau_2^2) &= \exp \left\{ -\frac{1}{2\tau_2^2} \boldsymbol{\eta}' \mathbf{R}_3^{-1} \boldsymbol{\eta} \right. \\ &\quad \left. + \sum_{i=1}^n \left( z_i f(x_i)' \boldsymbol{\eta} - \log(1 + e^{f(x_i)' \boldsymbol{\eta}}) \right) \right\}. \end{aligned}$$

The Metropolis-Hasting (MH) algorithm is the go-to method to sample from a target distribution with

an unknown distribution family. It generates a candidate sample from the proposal distribution, computes the acceptance probability, and then either accepts or rejects the candidate value based on the acceptance probability. More details can be found in Chapter 11 of Gelman et al. (2014). In the case of a logistic model for the binary response  $Z$ , Holmes and Held (2006) introduced a sampling method from the exact full-conditional distribution in Eq. [5] via auxiliary variables. We have implemented both algorithms and conducted some simulation examples, which are omitted here. Compared with the MH algorithm the Holmes and Held (2006) algorithm is not as efficient when embedded in the Gibbs sampling algorithm in Section 3.2. We thus choose the MH algorithm to sample from Eq. [5], but readers can choose whichever one of the two seems to fit their application.

The full-conditional distributions of  $\tau_1^2$  and  $\tau_2^2$  should follow an inverse- $\chi^2$  distribution due to the semiconjugacy; they are

$$\tau_1^2 | \text{rest}, \mathbf{y}, \mathbf{z} \sim \text{Inv} - \chi^2 \left( \nu + 2p, \frac{1}{\nu + 2p} [\boldsymbol{\beta}^{(1)'} \mathbf{R}_1^{-1} \boldsymbol{\beta}^{(1)} + \boldsymbol{\beta}^{(2)'} \mathbf{R}_2^{-1} \boldsymbol{\beta}^{(2)} + \nu \delta^2] \right), \quad [6]$$

$$\tau_2^2 | \text{rest}, \mathbf{y}, \mathbf{z} \sim \text{Inv} - \chi^2 \left( \nu + p, \frac{1}{\nu + p} [\boldsymbol{\eta}' \mathbf{R}_3^{-1} \boldsymbol{\eta} + \nu \delta^2] \right). \quad [7]$$

The exact full-conditional distributions of  $r_1$ ,  $r_2$ , and  $r_3$  are listed below:

$$\begin{aligned} p(r_1 | \text{rest parameters}, \mathbf{y}, \mathbf{z}) &\propto |\mathbf{R}_1|^{-\frac{1}{2}} \\ &\exp \left\{ -\frac{1}{2\tau_1^2} \boldsymbol{\beta}^{(1)'} \mathbf{R}_1^{-1} \boldsymbol{\beta}^{(1)} \right\} r_1^{a-1} (1-r_1)^{b-1}, \end{aligned} \quad [8]$$

$$\begin{aligned} p(r_2 | \text{rest parameters}, \mathbf{y}, \mathbf{z}) &\propto |\mathbf{R}_2|^{-\frac{1}{2}} \\ &\exp \left\{ -\frac{1}{2\tau_2^2} \boldsymbol{\beta}^{(2)'} \mathbf{R}_2^{-1} \boldsymbol{\beta}^{(2)} \right\} r_2^{a-1} (1-r_2)^{b-1}, \end{aligned} \quad [9]$$

$$\begin{aligned} p(r_3 | \text{rest parameters}, \mathbf{y}, \mathbf{z}) &\propto |\mathbf{R}_3|^{-\frac{1}{2}} \\ &\exp \left\{ -\frac{1}{2\tau_2^2} \boldsymbol{\eta}' \mathbf{R}_3^{-1} \boldsymbol{\eta} \right\} r_3^{a-1} (1-r_3)^{b-1}. \end{aligned} \quad [10]$$

Although the full-conditional distributions in Eqs. [8] through [10] are straightforward to derive, they are not any known distributions. Again, we use the MH algorithm to sample from them.

To sum up this section, the merits of the proposed Bayesian framework, besides overcoming the limitations of the frequentist approach, lie in the hierarchical structure. The hyperparameters  $(r_1, r_2, r_3)$  connect the coefficient parameters  $\beta^{(1)}$ ,  $\beta^{(2)}$ , and  $\boldsymbol{\eta}$  by following the same marginal hyperprior. In the same way the hyperparameters  $(\tau_1^2, \tau_2^2)$  connect the models for

Y and Z. The hierarchical model is flexible enough to provide a reasonable fit to the data—as shown later—and such flexibility is not paid by the price of too many parameters.

### 3.2. Gibbs sampling algorithm

For the blocks of parameters  $\beta^{(1)}$ ,  $\beta^{(2)}$ ,  $\sigma^2$ ,  $\tau^2_1$ , and  $\tau^2_2$  we can easily obtain their samples from the normal or inverse- $\chi^2$  distributions. For  $\eta$ ,  $r_1$ ,  $r_2$ , and  $r_3$  we use the MH algorithm since their full-conditional distributions are not from any known distribution family. Denote  $B$  and  $b$  as the length of the entire MCMC chain and the burn-in sample size, respectively. The following is the Gibbs sampling algorithm for the proposed Bayesian hierarchical model.

Step 1: Set the initial values for the hyperparameters  $r_{1,0}$ ,  $r_{2,0}$ ,  $r_{3,0}$ ,  $\tau^2_{1,0}$ , and  $\tau^2_{2,0}$ .

Step 2: Set the initial values for the parameters  $\beta_0^{(1)}$ ,  $\beta_0^{(2)}$ ,  $\eta_0$ , and  $\sigma_0^2$  as follows.

- Use the least squares estimates for  $\beta_0^{(1)}$ ,  $\beta_0^{(2)}$ . When the least squares estimates are not available choose the ridge regression estimates.
- Use the logistic regression estimate for  $\eta_0$ .
- Set  $\sigma_0^2 = \frac{1}{n}[(y - V_1 F \beta^{(1)} - V_2 F \beta^{(2)})^T (y - V_1 F \beta^{(1)} - V_2 F \beta^{(2)})]$ .

Step 3: For  $t = 1$  to  $B$ , do the following:

- Sample  $(\beta_t^{(1)}, \beta_t^{(2)})$  from  $N(\mu_\beta, \Sigma_\beta)$ .
- Sample  $\sigma_t^2$  from  $\text{Inv} - \chi^2(n + 0.001, \frac{\sum_{i=1}^n (y_i - \mu_i)^2 + 10^{-6}}{n + 0.001})$ , where  $\mu_i$ 's are the elements of  $V_1 F \beta_t^{(1)} + V_2 F \beta_t^{(2)}$ .
- Sample  $\eta_t$  by the MH algorithm from the distribution in Eq. [5].
- Sample  $\tau^2_{1,t}$  and  $\tau^2_{2,t}$  from Eqs. [6] and [7].
- Sample  $r_{1,t}$ ,  $r_{2,t}$  and  $r_{3,t}$  by the MH algorithm from the distributions in Eqs. [8] through [10].
- If the prediction at any query point  $x$  is required, sample  $z_t$  from the sampling distribution of  $Z|\eta_t$  in Eq. [1] and  $y_t$  from the sampling distribution of  $Y|z_t, \beta_t^{(1)}, \beta_t^{(2)}, \sigma_t^2$  in Eq. [2].

Step 4: Discard the first  $b$  samples  $(\theta_t, \phi_t)$  and posterior predictive samples  $(y_t, z_t)$  from the burn-in period. Using the rest of the samples, return the empirical posterior distributions for  $(\theta, \phi)$  and the empirical posterior predictive distributions for both responses at  $x$ .

There are three kinds of tuning parameters. The tuning parameters for the MCMC are  $(B, b)$ , which

can be set by considering the amount of computation needed to achieve the convergence and the preset accuracy level. The tuning parameters for the algorithm are the initial values  $(r_{1,0}, r_{2,0}, r_{3,0}, \tau^2_{1,0}, \tau^2_{2,0})$ . The tuning parameters for the Bayesian model are the parameters of the hyperprior distribution  $(\nu, \delta^2, a, b)$ , which can be chosen by analyzing historical data if they are available. Alternatively, we suggest setting  $(a, b)$  with  $\frac{a}{a+b} = 1/3$  or  $1/2$  and setting  $\nu = 2$  or  $\nu = 4.5$ , which lead to weakly informative distributions. Through a simulation-based sensitivity study (omitted from this article), we find that the Bayesian hierarchical model is robust to the choice of  $(\nu, \delta^2, a, b)$  around the suggested settings.

## 4. Simulation examples

### 4.1. A small-scale example

We first use a small-scale simulated example to demonstrate the estimation accuracy of the proposed model and the sampling algorithm. In this example we set the dimension of  $x$  as  $p = 5$  and the sample size  $n = 100$ ; we then generate the training and testing data from the joint QQ model specified in Eq. [3]. The input variable values  $x_1, \dots, x_n$  are independent and identically distributed samples from  $N(0, \Sigma)$ , where  $\Sigma = (\Sigma_{ij})_{p \times p}$  and the  $(i, j)$ th entry  $\Sigma_{ij} = 0.5^{|i-j|}$ . The true values of  $\beta^{(1)}$ ,  $\beta^{(2)}$ ,  $\eta$  are plotted as the red vertical lines in the posterior histograms in Figures 2, 3, and 4. The true value of  $\sigma^2$  is set to be one, also shown in Figure 4. We can see that the modes of the posterior distribution of the parameters  $\theta$  match their respective true values quite closely.

Using the Gibbs sampling algorithm in Section 3.2 we set the model tuning parameters  $(\nu, \delta^2, a, b) = (4.5, 5, 0.1, 0.2)$  and the MCMC tuning parameters  $(B, b) = (10,000, 1,000)$ . In Step 1 we set the initial values of  $(r_{1,0}, r_{2,0}, r_{3,0}, \tau^2_{1,0}, \tau^2_{2,0}) = (0.3, 0.3, 0.3, 0.5, 0.5)$ . Gelman and Rubin's (1992) convergence diagnostic is frequently used for monitoring the convergence of the MCMC output, and it is implemented in the R package coda by Plummer et al. (2006). The multivariate scale reduction factor values returned by the `gelman.diag` function for  $\beta^{(1)}$ ,  $\beta^{(2)}$ , and  $\eta$  are 1, 1, and 1.03, respectively, and the univariate scale reduction factor for  $\sigma^2$  is also 1, indicating that the MCMC chains of all the parameters have converged within  $B$  iterations.

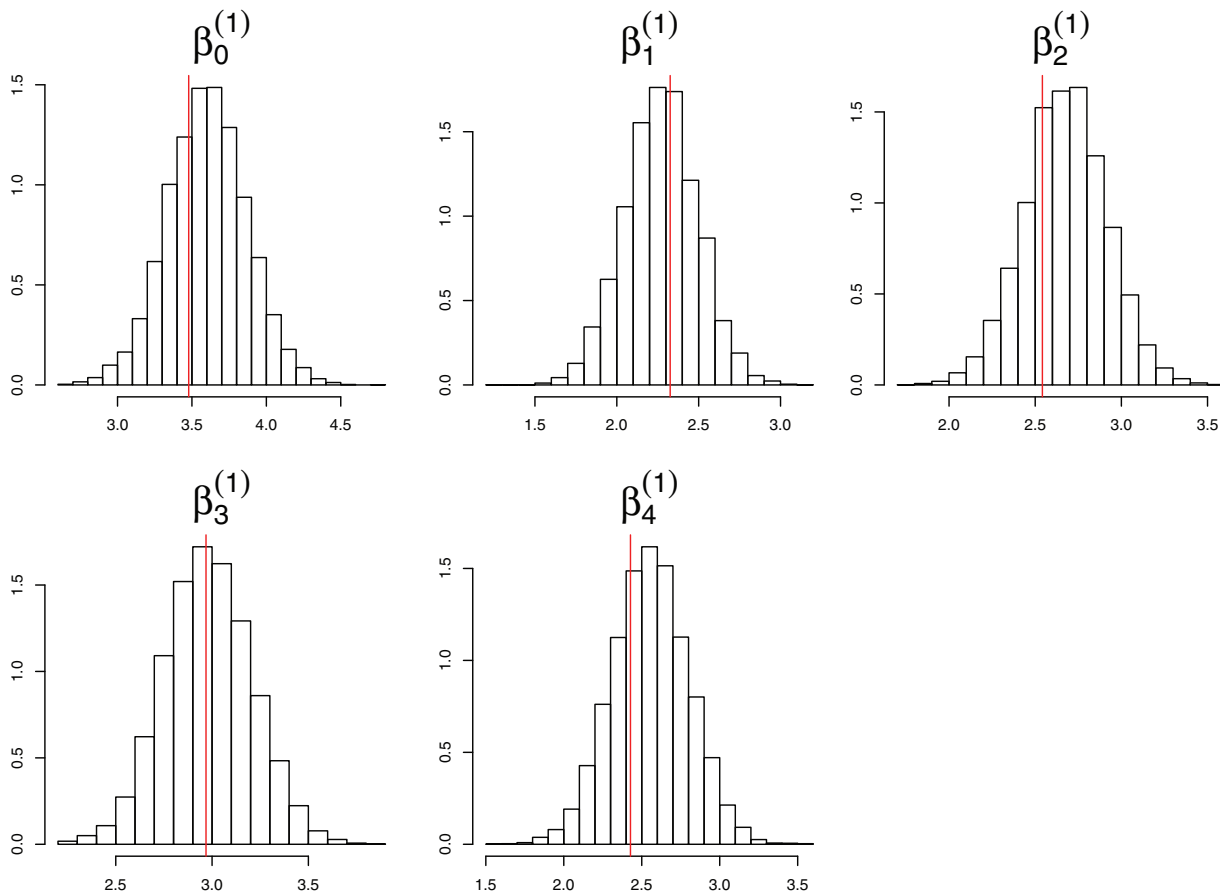


Figure 2. The histogram of the posterior samples of  $\beta^{(1)}$  compared against its true values plotted as the red vertical lines.

4.2. Comparisons with alternative methods

This subsection reports a large-scale simulation study on the proposed Bayesian hierarchical model and the posterior sampling scheme. We compare the following four methods.

- BHQQ(1): The proposed Bayesian hierarchical model and the model tuning parameters are  $(v, \delta^2, a, b) = (2, 2, 0.1, 0.1)$ .
- BHQQ(2): The proposed Bayesian hierarchical model and the model tuning parameters are  $(v, \delta^2, a, b) = (4.5, 5, 0.1, 0.2)$ .
- SM(F): Separate models of the two responses using a frequentist approach.
- SM(B): Separate models of the two responses using a Bayesian approach.

For both BHQQ(1) and (2) we set the length of the MCMC chain  $B = 10,000$  and burn-in size  $b = 1,000$  and let the initial value  $(r_{1,0}, r_{2,0}, r_{3,0}, \tau_{1,0}^2, \tau_{2,0}^2)$  be  $(0.3, 0.3, 0.3, 0.5, 0.5)$ . Both SM(F) and SM(B) fit a linear regression model for the continuous response  $Y$  and a logistic regression for the binary response  $Z$ . SM(F) employs the LASSO technique for both linear and

logistic regression models to shrink the estimated coefficients of parameters and select the significant variables. SM(B) sets the marginal normal priors for the parameters in linear and logistic regression models.

4.2.1. Different setups of data generation

We consider the similar simulation cases used in Deng and Jin (2015). There are two settings for the input dimension,  $p = 20$  and  $p = 50$ . The relatively high input-dimension settings are meant to test if the BHQQ can accurately identify the significant variables. To generate the data from the joint QQ model specified in Eq. [3], we consider different setups of the true values of  $\beta^{(1)}$  and  $\beta^{(2)}$ .

Let  $I_1$  and  $I_2$  be the index sets of the significant coefficients of  $\beta^{(1)}$  and  $\beta^{(2)}$ , respectively. Denote  $\bar{\beta}_1 = \{\beta_k^{(1)} : k \in I_1\}$  and  $\bar{\beta}_2 = \{\beta_k^{(2)} : k \in I_2\}$ . Similarly, denote  $I_3$  as the index set of significant coefficients of  $\eta$  and  $\bar{\eta} = \{\eta_k : k \in I_3\}$ . The coefficients that are not in  $I_1, I_2,$  and  $I_3$  are set to zero. The following four setups are used:

- Setup 1:  $I_1$  and  $I_2$  are the same, and the values of  $\bar{\beta}_1$  and  $\bar{\beta}_2$  are similar.
- Setup 2:  $I_1$  and  $I_2$  are different, but the values of  $\bar{\beta}_1$  and  $\bar{\beta}_2$  are similar.



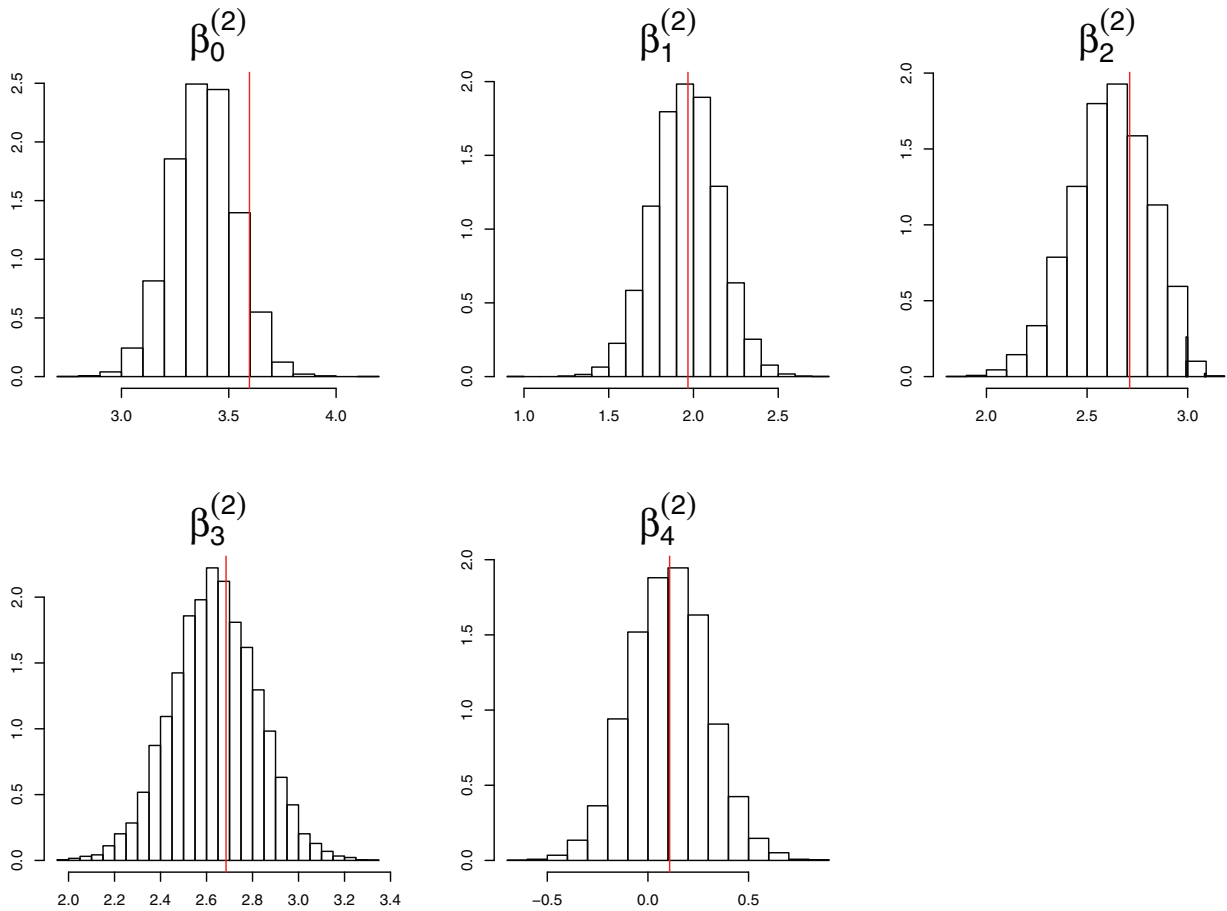


Figure 3. The histogram of the posterior samples of  $\beta^{(2)}$  compared against its true values plotted as the red vertical lines.

- Setup 3:  $I_1$  and  $I_2$  are the same, but the values of  $\bar{\beta}_1$  and  $\bar{\beta}_2$  are different.
- Setup 4:  $I_1$  and  $I_2$  are different, and the values of  $\bar{\beta}_1$  and  $\bar{\beta}_2$  are also different.

The four setups represent different relationships between  $Y|Z=0$  and  $Y|Z=1$ . Under Setup 1 the model of  $Y$  remains the same for different values of  $Z$  (i.e.,  $\beta^{(1)} = \beta^{(2)}$ ). Under Setups 2 - 4 there are different degrees of dependency of  $Y$  on  $Z$ , as  $\beta^{(1)}$  and  $\beta^{(2)}$  are different at different levels. Here we define *proportion sparsity*, denoted by  $s$ , as the proportion of non-zero entries in a parameter vector (i.e., the size of  $I_1$  or  $I_2$  divided by  $p$ ). In each Setup 1-4 we use two settings of  $s$ ,  $s = 0.2$  and  $s = 0.5$ . To sum up, we consider two settings of  $p$ , four settings of the differences between  $\beta^{(1)}$  and  $\beta^{(2)}$ , and two settings of the model sparsity  $s$ . Overall, the full factorial combinations have  $2 \times 4 \times 2 = 16$  cases.

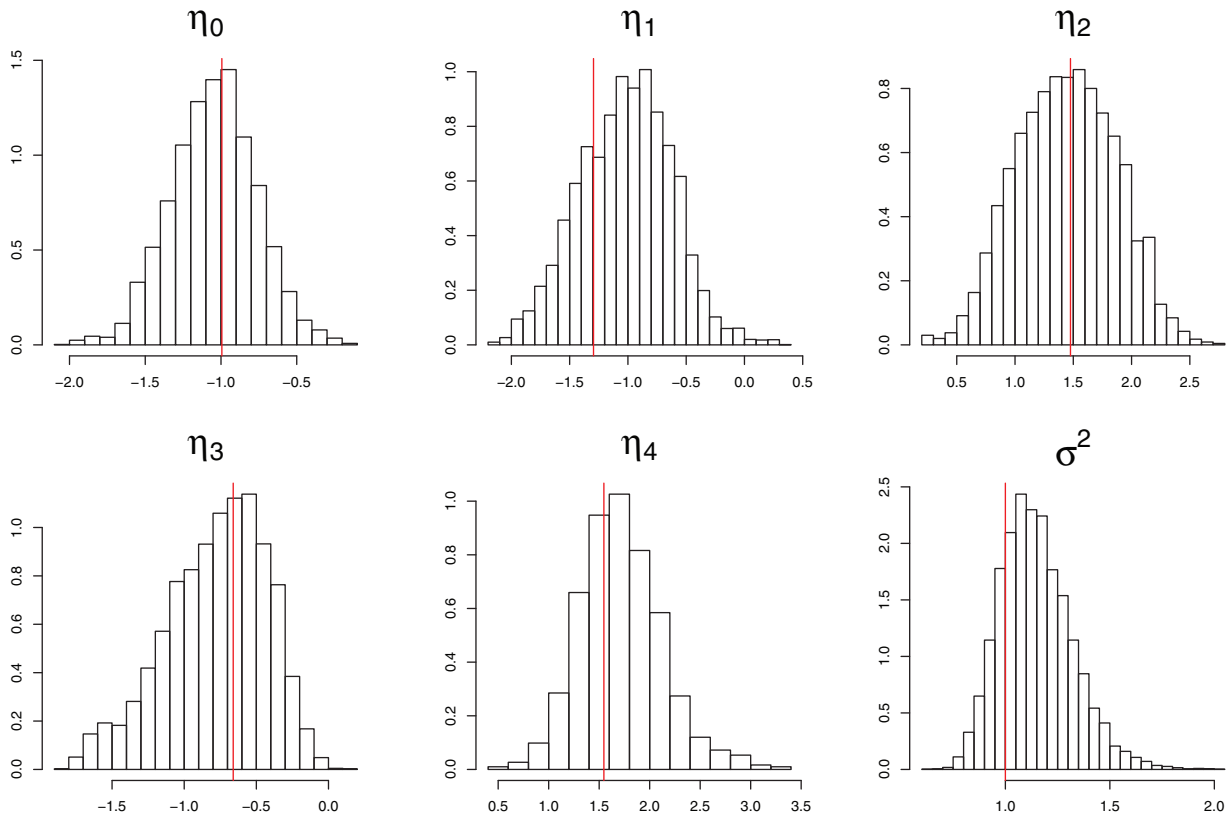
For each of the 16 cases the entry values of parameter vector  $\bar{\eta}$  are generated from the uniform distribution  $U(-2, 2)$ . For the cases under Setups 1 and 2 we first generate the parameter vector  $\bar{\beta}_1$  with each entry

value from the normal distribution  $N(2, 1)$  and then obtain the entry values of  $\bar{\beta}_2$  by adding a small perturbation from  $N(0, 0.01)$  to the values of  $\bar{\beta}_1$ . In Setups 3 and 4 the entry values of parameter vectors  $\bar{\beta}_1$  and  $\bar{\beta}_2$  are generated independently from  $N(2, 1)$ . The rest of the coefficients not in  $\bar{\eta}$ ,  $\bar{\beta}_1$ , and  $\bar{\beta}_2$  are zero. The  $n$  input design points  $x_1, \dots, x_n$  in the training set are independent and identically distributed (i.i.d.) samples from  $N(0, \Sigma)$ , where  $\Sigma = (\sigma_{ij})_{p \times p}$  with the  $(i, j)$ th entry  $\sigma_{ij} = 0.5^{|i-j|}$ . The  $n$  points of the testing data set are i.i.d. samples from  $U(-2, 2)$ . The sample size for both training and testing data sets is  $n = 100$ . The  $\sigma^2$  in Eq. [2] is one.

#### 4.2.2. Comparison criteria

We measure the performances of the four methods using three aspects: prediction accuracy, variable selection accuracy, and parameter estimation accuracy.

To measure the prediction accuracy we consider the root mean squared prediction error  $RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$  for the quantitative response  $Y$ , where  $\hat{y}_i$  represents the predicted value (the mean,



**Figure 4.** The histogram of the posterior samples of  $\boldsymbol{\eta}$  and  $\sigma^2$  compared against their true values plotted as the red vertical lines.

median, or mode of the posterior predictive distribution) of  $y_i$  in the testing data set. The prediction performance for the qualitative response  $Z$  is measured by the misclassification error  $ME = \frac{1}{n} \sum_{i=1}^n I(z_i \neq \hat{z}_i)$ , where  $\hat{z}_i$  is the predicted value of  $z_i$  and  $I(\cdot)$  is an indicator function. The prediction value  $\hat{z}_i$  is set as one if the corresponding posterior predictive sample mean or median of  $\Pr(Z = 1 | \mathbf{x}, \boldsymbol{\eta}_t)$  ( $\boldsymbol{\eta}_t$  is the MCMC chain) is larger than 0.5.

In addition, to gauge the performance of variable selection, false positive rates (FP) and false negative rates (FN) are defined. In the FP case the model estimation classifies insignificant predictors as significant ones. Similarly, FN estimation classifies significant predictors as insignificant ones. Here we use the total number of falsely selected variables  $\gamma = FP + FN$  as the performance measure. For the frequentist approach SM(F) the selected variables are from LASSO. For the Bayesian approach BHQQ and SM(B) the significant variables are selected based on the credit intervals from the rest of the 9,000 after burn-in. Denote  $\gamma_{lm}$  as the number of false selection for  $\boldsymbol{\beta}^{(1)}$  and  $\boldsymbol{\beta}^{(2)}$ , and  $\gamma_{logit}$  as the number of false selection for  $\boldsymbol{\eta}$ .

Moreover, to evaluate the parameter estimation accuracy,  $L_2$  loss for each parameter estimate is

considered as follows:

$$L_2(\boldsymbol{\beta}^{(1)}) = \frac{1}{p} \|\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(1)}\|_2^2, \quad L_2(\boldsymbol{\beta}^{(2)}) = \frac{1}{p} \|\hat{\boldsymbol{\beta}}^{(2)} - \boldsymbol{\beta}^{(2)}\|_2^2,$$

$$L_2(\boldsymbol{\eta}) = \frac{1}{p} \|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|_2^2,$$

where  $\|\cdot\|_2$  stands for the vector  $L_2$  norm. Here  $\hat{\boldsymbol{\beta}}^{(1)}$ ,  $\hat{\boldsymbol{\beta}}^{(2)}$ , and  $\hat{\boldsymbol{\eta}}$  are the estimated parameter values that are compared against the corresponding true parameter values. For Bayesian models we use the modes of the posterior samples as the parameter estimates.

### 4.2.3. Simulation results

Under each of the 16 cases of data-generation setup we repeat  $M = 100$  times of simulation. In each simulation we generate the new training and testing data and then fit BHQQ(1), BHQQ(2), SM(F), and SM(B) to make predictions. Therefore, we can obtain  $M = 100$  values for each of the seven performance measures. Table 1 presents the average values and their corresponding standard deviations (in parentheses) of the aforementioned seven criteria for  $p = 20$ . Table 2 displays the results for  $p = 50$ , and its conclusions are similar to the ones for  $p = 20$ . So we focus

**Table 1.** The averages and standard errors (in parentheses) of seven loss measures for  $p = 20$ .

Setup	BHQ(1)							BHQ(2)							SM(F)							SM(B)						
	s = 0.2		s = 0.5		s = 0.2		s = 0.5		s = 0.2		s = 0.5		s = 0.2		s = 0.5		s = 0.2		s = 0.5		s = 0.2		s = 0.5					
	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE				
1	RMSPE	0.664 (0.022)	0.910 (0.021)	0.656 (0.022)	0.908 (0.021)	0.464 (0.015)	0.547 (0.013)	0.576 (0.016)	0.664 (0.022)	0.910 (0.021)	0.656 (0.022)	0.908 (0.021)	0.464 (0.015)	0.547 (0.013)	0.576 (0.016)	0.664 (0.022)	0.910 (0.021)	0.656 (0.022)	0.908 (0.021)	0.464 (0.015)	0.547 (0.013)	0.576 (0.016)	0.664 (0.022)	0.910 (0.021)				
	ME	0.220 (0.006)	0.198 (0.005)	0.224 (0.006)	0.199 (0.005)	0.305 (0.006)	0.234 (0.005)	0.236 (0.005)	0.220 (0.006)	0.198 (0.005)	0.224 (0.006)	0.199 (0.005)	0.305 (0.006)	0.234 (0.005)	0.236 (0.005)	0.220 (0.006)	0.198 (0.005)	0.224 (0.006)	0.199 (0.005)	0.305 (0.006)	0.234 (0.005)	0.236 (0.005)	0.220 (0.006)	0.198 (0.005)				
	$\gamma_{lm}$	1.660 (0.157)	1.880 (0.134)	1.620 (0.148)	1.900 (0.128)	10.08 (0.423)	4.860 (0.200)	2.740 (0.252)	1.660 (0.157)	1.880 (0.134)	1.620 (0.148)	1.900 (0.128)	10.08 (0.423)	4.860 (0.200)	2.740 (0.252)	1.660 (0.157)	1.880 (0.134)	1.620 (0.148)	1.900 (0.128)	10.08 (0.423)	4.860 (0.200)	2.740 (0.252)	1.660 (0.157)	1.880 (0.134)				
	$\gamma_{logit}$	1.810 (0.104)	4.710 (0.121)	1.770 (0.094)	4.790 (0.123)	5.930 (0.178)	6.080 (0.191)	1.720 (0.087)	1.810 (0.104)	4.710 (0.121)	1.770 (0.094)	4.790 (0.123)	5.930 (0.178)	6.080 (0.191)	1.720 (0.087)	1.810 (0.104)	4.710 (0.121)	1.770 (0.094)	4.790 (0.123)	5.930 (0.178)	6.080 (0.191)	1.720 (0.087)	1.810 (0.104)	4.710 (0.121)				
	$L2(\beta^{(1)})$	0.027 (0.003)	0.031 (0.002)	0.026 (0.003)	0.031 (0.002)	0.009 (0.001)	0.013 (0.001)	0.011 (0.001)	0.027 (0.003)	0.031 (0.002)	0.026 (0.003)	0.031 (0.002)	0.009 (0.001)	0.013 (0.001)	0.011 (0.001)	0.027 (0.003)	0.031 (0.002)	0.026 (0.003)	0.031 (0.002)	0.009 (0.001)	0.013 (0.001)	0.011 (0.001)	0.027 (0.003)	0.031 (0.002)				
	$L2(\beta^{(2)})$	0.025 (0.002)	0.069 (0.005)	0.025 (0.002)	0.068 (0.005)	0.009 (0.001)	0.013 (0.001)	0.011 (0.001)	0.025 (0.002)	0.069 (0.005)	0.025 (0.002)	0.068 (0.005)	0.009 (0.001)	0.013 (0.001)	0.011 (0.001)	0.025 (0.002)	0.069 (0.005)	0.025 (0.002)	0.068 (0.005)	0.009 (0.001)	0.013 (0.001)	0.011 (0.001)	0.025 (0.002)	0.069 (0.005)				
	$L2(\eta)$	0.075 (0.008)	0.288 (0.027)	0.081 (0.011)	0.374 (0.036)	0.060 (0.003)	0.254 (0.039)	0.064 (0.006)	0.075 (0.008)	0.288 (0.027)	0.081 (0.011)	0.374 (0.036)	0.060 (0.003)	0.254 (0.039)	0.064 (0.006)	0.075 (0.008)	0.288 (0.027)	0.081 (0.011)	0.374 (0.036)	0.060 (0.003)	0.254 (0.039)	0.064 (0.006)	0.075 (0.008)	0.288 (0.027)				
2	RMSPE	2.346 (0.030)	3.625 (0.065)	2.342 (0.030)	3.632 (0.068)	3.127 (0.030)	5.145 (0.045)	5.444 (0.050)	2.346 (0.030)	3.625 (0.065)	2.342 (0.030)	3.632 (0.068)	3.127 (0.030)	5.145 (0.045)	5.444 (0.050)	2.346 (0.030)	3.625 (0.065)	2.342 (0.030)	3.632 (0.068)	3.127 (0.030)	5.145 (0.045)	5.444 (0.050)	2.346 (0.030)	3.625 (0.065)				
	ME	0.252 (0.006)	0.196 (0.005)	0.247 (0.005)	0.196 (0.005)	0.326 (0.006)	0.234 (0.005)	0.297 (0.005)	0.252 (0.006)	0.196 (0.005)	0.247 (0.005)	0.196 (0.005)	0.326 (0.006)	0.234 (0.005)	0.297 (0.005)	0.252 (0.006)	0.196 (0.005)	0.247 (0.005)	0.196 (0.005)	0.326 (0.006)	0.234 (0.005)	0.297 (0.005)	0.252 (0.006)	0.196 (0.005)				
	$\gamma_{lm}$	1.970 (0.170)	2.480 (0.140)	1.900 (0.176)	2.550 (0.147)	17.68 (0.552)	16.52 (0.266)	12.08 (0.175)	1.970 (0.170)	2.480 (0.140)	1.900 (0.176)	2.550 (0.147)	17.68 (0.552)	16.52 (0.266)	12.08 (0.175)	1.970 (0.170)	2.480 (0.140)	1.900 (0.176)	2.550 (0.147)	17.68 (0.552)	16.52 (0.266)	12.08 (0.175)	1.970 (0.170)	2.480 (0.140)				
	$\gamma_{logit}$	2.410 (0.112)	4.920 (0.139)	2.340 (0.097)	5.070 (0.165)	5.830 (0.278)	6.510 (0.180)	5.320 (0.146)	2.410 (0.112)	4.920 (0.139)	2.340 (0.097)	5.070 (0.165)	5.830 (0.278)	6.510 (0.180)	5.320 (0.146)	2.410 (0.112)	4.920 (0.139)	2.340 (0.097)	5.070 (0.165)	5.830 (0.278)	6.510 (0.180)	5.320 (0.146)	2.410 (0.112)	4.920 (0.139)				
	$L2(\beta^{(1)})$	0.024 (0.002)	0.046 (0.003)	0.023 (0.002)	0.046 (0.003)	0.366 (0.010)	1.066 (0.029)	1.023 (0.039)	0.024 (0.002)	0.046 (0.003)	0.023 (0.002)	0.046 (0.003)	0.366 (0.010)	1.066 (0.029)	1.023 (0.039)	0.024 (0.002)	0.046 (0.003)	0.023 (0.002)	0.046 (0.003)	0.366 (0.010)	1.066 (0.029)	1.023 (0.039)	0.024 (0.002)	0.046 (0.003)				
	$L2(\beta^{(2)})$	0.028 (0.002)	0.050 (0.003)	0.028 (0.002)	0.050 (0.003)	0.429 (0.012)	1.069 (0.032)	1.222 (0.041)	0.028 (0.002)	0.050 (0.003)	0.028 (0.002)	0.050 (0.003)	0.429 (0.012)	1.069 (0.032)	1.222 (0.041)	0.028 (0.002)	0.050 (0.003)	0.028 (0.002)	0.050 (0.003)	0.429 (0.012)	1.069 (0.032)	1.222 (0.041)	0.028 (0.002)	0.050 (0.003)				
	$L2(\eta)$	0.078 (0.004)	0.291 (0.019)	0.083 (0.006)	0.410 (0.037)	0.061 (0.003)	0.192 (0.008)	0.218 (0.009)	0.078 (0.004)	0.291 (0.019)	0.083 (0.006)	0.410 (0.037)	0.061 (0.003)	0.192 (0.008)	0.218 (0.009)	0.078 (0.004)	0.291 (0.019)	0.083 (0.006)	0.410 (0.037)	0.061 (0.003)	0.192 (0.008)	0.218 (0.009)	0.078 (0.004)	0.291 (0.019)				
3	RMSPE	1.196 (0.017)	1.455 (0.021)	1.191 (0.017)	1.481 (0.021)	1.416 (0.012)	1.724 (0.017)	1.746 (0.019)	1.196 (0.017)	1.455 (0.021)	1.191 (0.017)	1.481 (0.021)	1.416 (0.012)	1.724 (0.017)	1.746 (0.019)	1.196 (0.017)	1.455 (0.021)	1.191 (0.017)	1.481 (0.021)	1.416 (0.012)	1.724 (0.017)	1.746 (0.019)	1.196 (0.017)	1.455 (0.021)				
	ME	0.202 (0.005)	0.204 (0.005)	0.202 (0.005)	0.209 (0.005)	0.277 (0.006)	0.254 (0.005)	0.236 (0.005)	0.202 (0.005)	0.204 (0.005)	0.202 (0.005)	0.209 (0.005)	0.277 (0.006)	0.254 (0.005)	0.236 (0.005)	0.202 (0.005)	0.204 (0.005)	0.202 (0.005)	0.209 (0.005)	0.277 (0.006)	0.254 (0.005)	0.236 (0.005)	0.202 (0.005)	0.204 (0.005)				
	$\gamma_{lm}$	1.940 (0.161)	2.800 (0.127)	1.960 (0.161)	2.830 (0.132)	12.30 (0.570)	11.20 (0.461)	2.540 (0.200)	1.940 (0.161)	2.800 (0.127)	1.960 (0.161)	2.830 (0.132)	12.30 (0.570)	11.20 (0.461)	2.540 (0.200)	1.940 (0.161)	2.800 (0.127)	1.960 (0.161)	2.830 (0.132)	12.30 (0.570)	11.20 (0.461)	2.540 (0.200)	1.940 (0.161)	2.800 (0.127)				
	$\gamma_{logit}$	1.300 (0.111)	5.670 (0.153)	1.320 (0.113)	5.810 (0.136)	4.480 (0.243)	6.270 (0.224)	5.660 (0.109)	1.300 (0.111)	5.670 (0.153)	1.320 (0.113)	5.810 (0.136)	4.480 (0.243)	6.270 (0.224)	5.660 (0.109)	1.300 (0.111)	5.670 (0.153)	1.320 (0.113)	5.810 (0.136)	4.480 (0.243)	6.270 (0.224)	5.660 (0.109)	1.300 (0.111)	5.670 (0.153)				
	$L2(\beta^{(1)})$	0.016 (0.002)	0.021 (0.001)	0.016 (0.002)	0.021 (0.001)	0.062 (0.002)	0.100 (0.003)	0.094 (0.004)	0.016 (0.002)	0.021 (0.001)	0.016 (0.002)	0.021 (0.001)	0.062 (0.002)	0.100 (0.003)	0.094 (0.004)	0.016 (0.002)	0.021 (0.001)	0.016 (0.002)	0.021 (0.001)	0.062 (0.002)	0.100 (0.003)	0.094 (0.004)	0.016 (0.002)	0.021 (0.001)				
	$L2(\beta^{(2)})$	0.046 (0.004)	0.105 (0.010)	0.047 (0.004)	0.107 (0.010)	0.102 (0.003)	0.185 (0.003)	0.198 (0.005)	0.046 (0.004)	0.105 (0.010)	0.047 (0.004)	0.107 (0.010)	0.102 (0.003)	0.185 (0.003)	0.198 (0.005)	0.046 (0.004)	0.105 (0.010)	0.047 (0.004)	0.107 (0.010)	0.102 (0.003)	0.185 (0.003)	0.198 (0.005)	0.046 (0.004)	0.105 (0.010)				
	$L2(\eta)$	0.108 (0.009)	0.180 (0.017)	0.126 (0.012)	0.205 (0.020)	0.073 (0.004)	0.129 (0.006)	0.141 (0.008)	0.108 (0.009)	0.180 (0.017)	0.126 (0.012)	0.205 (0.020)	0.073 (0.004)	0.129 (0.006)	0.141 (0.008)	0.108 (0.009)	0.180 (0.017)	0.126 (0.012)	0.205 (0.020)	0.073 (0.004)	0.129 (0.006)	0.141 (0.008)	0.108 (0.009)	0.180 (0.017)				
4	RMSPE	1.699 (0.024)	2.955 (0.048)	1.714 (0.026)	3.006 (0.057)	2.425 (0.025)	4.210 (0.035)	4.526 (0.044)	1.699 (0.024)	2.955 (0.048)	1.714 (0.026)	3.006 (0.057)	2.425 (0.025)	4.210 (0.035)	4.526 (0.044)	1.699 (0.024)	2.955 (0.048)	1.714 (0.026)	3.006 (0.057)	2.425 (0.025)	4.210 (0.035)	4.526 (0.044)	1.699 (0.024)	2.955 (0.048)				
	ME	0.201 (0.005)	0.216 (0.005)	0.201 (0.006)	0.223 (0.005)	0.286 (0.005)	0.255 (0.005)	0.257 (0.006)	0.201 (0.005)	0.216 (0.005)	0.201 (0.006)	0.223 (0.005)	0.286 (0.005)	0.255 (0.005)	0.257 (0.006)	0.201 (0.005)	0.216 (0.005)	0.201 (0.006)	0.223 (0.005)	0.286 (0.005)	0.255 (0.005)	0.257 (0.006)	0.201 (0.005)	0.216 (0.005)				
	$\gamma_{lm}$	1.680 (0.153)	2.370 (0.121)	1.760 (0.164)	2.340 (0.117)	17.86 (0.546)	14.78 (0.290)	9.580 (0.178)	1.680 (0.153)	2.370 (0.121)	1.760 (0.164)	2.340 (0.117)	17.86 (0.546)	14.78 (0.290)	9.580 (0.178)	1.680 (0.153)	2.370 (0.121)	1.760 (0.164)	2.340 (0.117)	17.86 (0.546)	14.78 (0.290)	9.580 (0.178)	1.680 (0.153)	2.370 (0.121)				
	$\gamma_{logit}$	2.490 (0.100)	4.350 (0.134)	2.380 (0.094)	4.700 (0.149)	5.630 (0.276)	5.330 (0.181)	4.650 (0.118)	2.490 (0.100)	4.350 (0.134)	2.380 (0.094)	4.700 (0.149)	5.630 (0.276)	5.330 (0.181)	4.650 (0.118)	2.490 (0.100)	4.350 (0.134)	2.380 (0.094)	4.700 (0.149)	5.630 (0.276)	5.330 (0.181)	4.650 (0.118)	2.490 (0.100)	4.350 (0.134)				
	$L2(\beta^{(1)})$	0.029 (0.003)	0.045 (0.003)	0.029 (0.003)	0.045 (0.003)	0.278 (0.008)	0.662 (0.020)	0.763 (0.027)	0.029 (0.003)	0.045 (0.003)	0.029 (0.003)	0.045 (0.003)	0.278 (0.008)	0.662 (0.020)	0.763 (0.027)	0.029 (0.003)	0.045 (0.003)	0.029 (0.003)	0.045 (0.003)	0.278 (0.008)	0.662 (0.020)	0.763 (0.027)	0.029 (0.003)	0.045 (0.003)				
	$L2(\beta^{(2)})$	0.019 (0.002)	0.034 (0.002)	0.020 (0.002)	0.034 (0.002)	0.182 (0.007)	0.795 (0.021)	0.893 (0.026)	0.019 (0.002)	0.034 (0.002)	0.020 (0.002)	0.034 (0.002)	0.182 (0.007)	0.795 (0.021)	0.893 (0.026)	0.019 (0.002)	0.034 (0.002)	0.020 (0.002)	0.034 (0.002)	0.182 (0.007)	0.795 (0.021)	0.893 (0.026)	0.019 (0.002)	0.034 (0.002)				
	$L2(\eta)$	0.065 (0.005)	0.268 (0.039)	0.077 (0.007)	0.329 (0.036)	0.054 (0.003)	0.160 (0.007)	0.176 (0.009)	0.065 (0.005)	0.268 (0.039)	0.077 (0.007)	0.329 (0.036)	0.054 (0.003)	0.160 (0.007)	0.176 (0.009)	0.065 (0.005)	0.268 (0.039)	0.077 (0.007)	0.329 (0.036)	0.054 (0.003)	0.160 (0.007)	0.176 (0.009)	0.065 (0.005)	0.268 (0.039)				

Table 2. The averages and standard errors (in parentheses) of seven loss measures for  $p = 50$ .

Setup	BHQO(1)							BHQO(2)							SM(F)							SM(B)																																					
	$s = 0.2$		$s = 0.5$		$s = 0.2$		$s = 0.5$		$s = 0.2$		$s = 0.5$		$s = 0.2$		$s = 0.5$		$s = 0.2$		$s = 0.5$		$s = 0.2$		$s = 0.5$		$s = 0.2$		$s = 0.5$																																
1	RMSPE	2.227 (0.053)	3.841 (0.083)	2.188 (0.054)	3.795 (0.079)	0.741 (0.017)	1.130 (0.021)	1.174 (0.027)	1.326 (0.026)	0.250 (0.007)	0.305 (0.006)	0.255 (0.007)	0.314 (0.005)	0.288 (0.005)	0.315 (0.006)	0.308 (0.007)	0.315 (0.006)	0.308 (0.007)	0.315 (0.006)	6.120 (0.268)	15.49 (0.379)	6.080 (0.267)	15.61 (0.397)	27.90 (0.997)	29.64 (0.848)	13.88 (0.679)	13.88 (0.679)	7.540 (0.113)	22.99 (0.119)	7.540 (0.113)	22.99 (0.119)	0.098 (0.006)	0.140 (0.009)	0.098 (0.007)	0.140 (0.009)	0.009 (0.001)	0.020 (0.001)	0.022 (0.001)	0.027 (0.001)	0.094 (0.006)	0.511 (0.030)	0.093 (0.006)	0.524 (0.032)	0.009 (0.001)	0.020 (0.001)	0.022 (0.001)	0.027 (0.001)	0.152 (0.006)	0.417 (0.009)	0.174 (0.007)	0.429 (0.008)	0.133 (0.004)	0.348 (0.007)	0.168 (0.005)	0.481 (0.007)				
	ME	0.270 (0.007)	0.304 (0.005)	0.269 (0.007)	0.304 (0.006)	0.292 (0.006)	0.290 (0.008)	0.293 (0.008)	0.366 (0.008)	7.310 (0.140)	19.48 (0.206)	7.270 (0.144)	19.77 (0.211)	14.14 (0.393)	15.93 (0.290)	7.360 (0.349)	7.360 (0.349)	26.42 (0.558)	26.42 (0.558)	0.095 (0.005)	0.868 (0.061)	0.095 (0.005)	0.854 (0.061)	0.510 (0.013)	2.020 (0.039)	0.763 (0.021)	2.946 (0.066)	0.098 (0.007)	0.208 (0.019)	0.099 (0.007)	0.207 (0.019)	0.511 (0.014)	1.804 (0.042)	0.762 (0.025)	2.585 (0.064)	0.154 (0.010)	0.448 (0.019)	0.132 (0.006)	0.474 (0.012)	0.096 (0.003)	0.384 (0.008)	0.119 (0.005)	0.505 (0.013)																
	$\gamma_{lm}$	7.760 (0.299)	14.43 (0.420)	7.910 (0.312)	14.35 (0.411)	37.90 (0.749)	41.06 (0.424)	26.42 (0.558)	40.20 (0.461)	7.280 (0.135)	18.94 (0.241)	7.180 (0.137)	19.25 (0.249)	12.78 (0.391)	16.45 (0.317)	6.910 (0.121)	6.910 (0.121)	15.50 (0.664)	15.50 (0.664)	0.034 (0.002)	0.304 (0.023)	0.034 (0.003)	0.308 (0.023)	0.091 (0.003)	0.474 (0.014)	0.162 (0.006)	0.618 (0.019)	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)																
	$\gamma_{logit}$	0.034 (0.002)	0.304 (0.023)	0.034 (0.003)	0.308 (0.023)	0.091 (0.003)	0.474 (0.014)	0.162 (0.006)	0.618 (0.019)	0.034 (0.002)	0.304 (0.023)	0.034 (0.003)	0.308 (0.023)	0.091 (0.003)	0.474 (0.014)	0.162 (0.006)	0.618 (0.019)	0.034 (0.002)	0.304 (0.023)	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)																								
	$L2(\beta^{(1)})$	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)	0.158 (0.008)	0.555 (0.012)	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)																								
	$L2(\beta^{(2)})$	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)	4.660 (0.056)	7.482 (0.095)	4.637 (0.059)	7.469 (0.098)	6.162 (0.048)	8.813 (0.092)	7.781 (0.101)	10.60 (0.122)	4.660 (0.056)	7.482 (0.095)	4.637 (0.059)	7.469 (0.098)	6.162 (0.048)	8.813 (0.092)	7.781 (0.101)	10.60 (0.122)	0.256 (0.006)	0.311 (0.007)	0.252 (0.006)	0.307 (0.006)	0.306 (0.006)	0.293 (0.005)	0.299 (0.006)	0.380 (0.006)	0.540 (0.028)	18.24 (0.386)	0.540 (0.028)	18.24 (0.386)	40.38 (0.840)	39.92 (0.424)	29.10 (0.575)	36.86 (0.428)	0.080 (0.005)	0.989 (0.057)	0.080 (0.005)	1.026 (0.058)	0.553 (0.014)	1.630 (0.028)	0.923 (0.026)	2.141 (0.043)	0.094 (0.006)	0.099 (0.006)	0.089 (0.005)	0.099 (0.006)	0.548 (0.014)	0.884 (0.023)	0.864 (0.027)	1.431 (0.042)		
	$L2(\eta)$	0.116 (0.005)	0.502 (0.011)	0.124 (0.010)	0.509 (0.010)	0.086 (0.003)	0.423 (0.008)	0.113 (0.005)	0.582 (0.009)	0.116 (0.005)	0.502 (0.011)	0.124 (0.010)	0.509 (0.010)	0.086 (0.003)	0.423 (0.008)	0.113 (0.005)	0.582 (0.009)	0.116 (0.005)	0.502 (0.011)	0.094 (0.006)	0.099 (0.006)	0.089 (0.005)	0.099 (0.006)	0.548 (0.014)	0.884 (0.023)	0.864 (0.027)	1.431 (0.042)	0.094 (0.006)	0.099 (0.006)	0.089 (0.005)	0.099 (0.006)	0.548 (0.014)	0.884 (0.023)	0.864 (0.027)	1.431 (0.042)																								
2	RMSPE	4.555 (0.054)	8.801 (0.107)	4.549 (0.057)	8.844 (0.123)	5.827 (0.048)	10.83 (0.106)	7.096 (0.076)	13.23 (0.140)	0.270 (0.007)	0.304 (0.005)	0.269 (0.007)	0.304 (0.006)	0.292 (0.006)	0.290 (0.008)	0.293 (0.008)	0.366 (0.008)	0.293 (0.008)	0.366 (0.008)	7.760 (0.299)	14.43 (0.420)	7.910 (0.312)	14.35 (0.411)	37.90 (0.749)	41.06 (0.424)	26.42 (0.558)	40.20 (0.461)	7.310 (0.140)	19.48 (0.206)	7.270 (0.144)	19.77 (0.211)	14.14 (0.393)	15.93 (0.290)	7.360 (0.349)	23.50 (0.115)	0.095 (0.005)	0.868 (0.061)	0.095 (0.005)	0.854 (0.061)	0.510 (0.013)	2.020 (0.039)	0.763 (0.021)	2.946 (0.066)	0.098 (0.007)	0.208 (0.019)	0.099 (0.007)	0.207 (0.019)	0.511 (0.014)	1.804 (0.042)	0.762 (0.025)	2.585 (0.064)	0.154 (0.010)	0.448 (0.019)	0.132 (0.006)	0.474 (0.012)	0.096 (0.003)	0.384 (0.008)	0.119 (0.005)	0.505 (0.013)
	ME	0.270 (0.007)	0.304 (0.005)	0.269 (0.007)	0.304 (0.006)	0.292 (0.006)	0.290 (0.008)	0.293 (0.008)	0.366 (0.008)	7.760 (0.299)	14.43 (0.420)	7.910 (0.312)	14.35 (0.411)	37.90 (0.749)	41.06 (0.424)	26.42 (0.558)	40.20 (0.461)	7.760 (0.299)	14.43 (0.420)	7.310 (0.140)	19.48 (0.206)	7.270 (0.144)	19.77 (0.211)	14.14 (0.393)	15.93 (0.290)	7.360 (0.349)	23.50 (0.115)	0.095 (0.005)	0.868 (0.061)	0.095 (0.005)	0.854 (0.061)	0.510 (0.013)	2.020 (0.039)	0.763 (0.021)	2.946 (0.066)	0.098 (0.007)	0.208 (0.019)	0.099 (0.007)	0.207 (0.019)	0.511 (0.014)	1.804 (0.042)	0.762 (0.025)	2.585 (0.064)																
	$\gamma_{lm}$	7.760 (0.299)	14.43 (0.420)	7.910 (0.312)	14.35 (0.411)	37.90 (0.749)	41.06 (0.424)	26.42 (0.558)	40.20 (0.461)	7.310 (0.140)	19.48 (0.206)	7.270 (0.144)	19.77 (0.211)	14.14 (0.393)	15.93 (0.290)	7.360 (0.349)	23.50 (0.115)	7.760 (0.299)	14.43 (0.420)	7.310 (0.140)	19.48 (0.206)	7.270 (0.144)	19.77 (0.211)	14.14 (0.393)	15.93 (0.290)	7.360 (0.349)	23.50 (0.115)	0.095 (0.005)	0.868 (0.061)	0.095 (0.005)	0.854 (0.061)	0.510 (0.013)	2.020 (0.039)	0.763 (0.021)	2.946 (0.066)	0.098 (0.007)	0.208 (0.019)	0.099 (0.007)	0.207 (0.019)	0.511 (0.014)	1.804 (0.042)	0.762 (0.025)	2.585 (0.064)																
	$\gamma_{logit}$	0.034 (0.002)	0.304 (0.023)	0.034 (0.003)	0.308 (0.023)	0.091 (0.003)	0.474 (0.014)	0.162 (0.006)	0.618 (0.019)	0.034 (0.002)	0.304 (0.023)	0.034 (0.003)	0.308 (0.023)	0.091 (0.003)	0.474 (0.014)	0.162 (0.006)	0.618 (0.019)	0.034 (0.002)	0.304 (0.023)	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)																								
	$L2(\beta^{(1)})$	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)	0.158 (0.008)	0.555 (0.012)	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)																								
	$L2(\beta^{(2)})$	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)	4.660 (0.056)	7.482 (0.095)	4.637 (0.059)	7.469 (0.098)	6.162 (0.048)	8.813 (0.092)	7.781 (0.101)	10.60 (0.122)	4.660 (0.056)	7.482 (0.095)	4.637 (0.059)	7.469 (0.098)	6.162 (0.048)	8.813 (0.092)	7.781 (0.101)	10.60 (0.122)	0.256 (0.006)	0.311 (0.007)	0.252 (0.006)	0.307 (0.006)	0.306 (0.006)	0.293 (0.005)	0.299 (0.006)	0.380 (0.006)	0.540 (0.028)	18.24 (0.386)	0.540 (0.028)	18.24 (0.386)	40.38 (0.840)	39.92 (0.424)	29.10 (0.575)	36.86 (0.428)																		
	$L2(\eta)$	0.116 (0.005)	0.502 (0.011)	0.124 (0.010)	0.509 (0.010)	0.086 (0.003)	0.423 (0.008)	0.113 (0.005)	0.582 (0.009)	0.116 (0.005)	0.502 (0.011)	0.124 (0.010)	0.509 (0.010)	0.086 (0.003)	0.423 (0.008)	0.113 (0.005)	0.582 (0.009)	0.116 (0.005)	0.502 (0.011)	0.094 (0.006)	0.099 (0.006)	0.089 (0.005)	0.099 (0.006)	0.548 (0.014)	0.884 (0.023)	0.864 (0.027)	1.431 (0.042)	0.094 (0.006)	0.099 (0.006)	0.089 (0.005)	0.099 (0.006)	0.548 (0.014)	0.884 (0.023)	0.864 (0.027)	1.431 (0.042)																								
3	RMSPE	2.833 (0.040)	5.194 (0.077)	2.855 (0.037)	5.244 (0.071)	2.860 (0.031)	5.474 (0.067)	3.606 (0.053)	6.272 (0.083)	0.269 (0.008)	0.303 (0.006)	0.270 (0.007)	0.317 (0.006)	0.294 (0.006)	0.294 (0.006)	0.276 (0.008)	0.373 (0.008)	0.276 (0.008)	0.373 (0.008)	7.760 (0.282)	14.53 (0.367)	7.720 (0.263)	14.63 (0.368)	26.90 (1.065)	27.64 (0.731)	15.50 (0.664)	15.70 (0.574)	7.280 (0.135)	18.94 (0.241)	7.180 (0.137)	19.25 (0.249)	12.78 (0.391)	16.45 (0.317)	6.910 (0.121)	23.26 (0.124)	0.034 (0.002)	0.304 (0.023)	0.034 (0.003)	0.308 (0.023)	0.091 (0.003)	0.474 (0.014)	0.162 (0.006)	0.618 (0.019)	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)	0.158 (0.008)	0.555 (0.012)	0.171 (0.011)	0.583 (0.013)	0.103 (0.004)	0.475 (0.009)	0.120 (0.006)	0.631 (0.014)
	ME	0.269 (0.008)	0.303 (0.006)	0.270 (0.007)	0.317 (0.006)	0.294 (0.006)	0.294 (0.006)	0.276 (0.008)	0.373 (0.008)	7.760 (0.282)	14.53 (0.367)	7.720 (0.263)	14.63 (0.368)	26.90 (1.065)	27.64 (0.731)	15.50 (0.664)	15.70 (0.574)	7.760 (0.282)	14.53 (0.367)	7.280 (0.135)	18.94 (0.241)	7.180 (0.137)	19.25 (0.249)	12.78 (0.391)	16.45 (0.317)	6.910 (0.121)	23.26 (0.124)	0.034 (0.002)	0.304 (0.023)	0.034 (0.003)	0.308 (0.023)	0.091 (0.003)	0.474 (0.014)	0.162 (0.006)	0.618 (0.019)	0.230 (0.014)	0.312 (0.021)	0.232 (0.014)	0.311 (0.020)	0.181 (0.005)	0.450 (0.012)	0.265 (0.007)	0.619 (0.016)																
	$\gamma_{lm}$	7.760 (0.282)	14.53 (0.367)	7.720 (0.263)	14.63 (0.368)	26.90 (1.065)	27.64 (0.731)	15.50 (0.664)	15.70 (0.574)	7.280 (0.135)	18.94 (0.241)	7.180 (0.137)	19.25 (0.249)	12.78 (0.391)	16.45 (0.317)	6.910 (0.121)	23.26 (0.124)	7.760 (0.282)	14.53 (0.367)	7.280 (0.135)	18.94 (0.241)	7.180 (0.137)	19.25 (0.249)	12.78 (0.391)	16.45 (0.317)	6.910 (0.121)	23.26 (0.124)	0.034 (0.002)	0.304 (0.023)	0.034 (0.003)	0.308 (0.023)																												

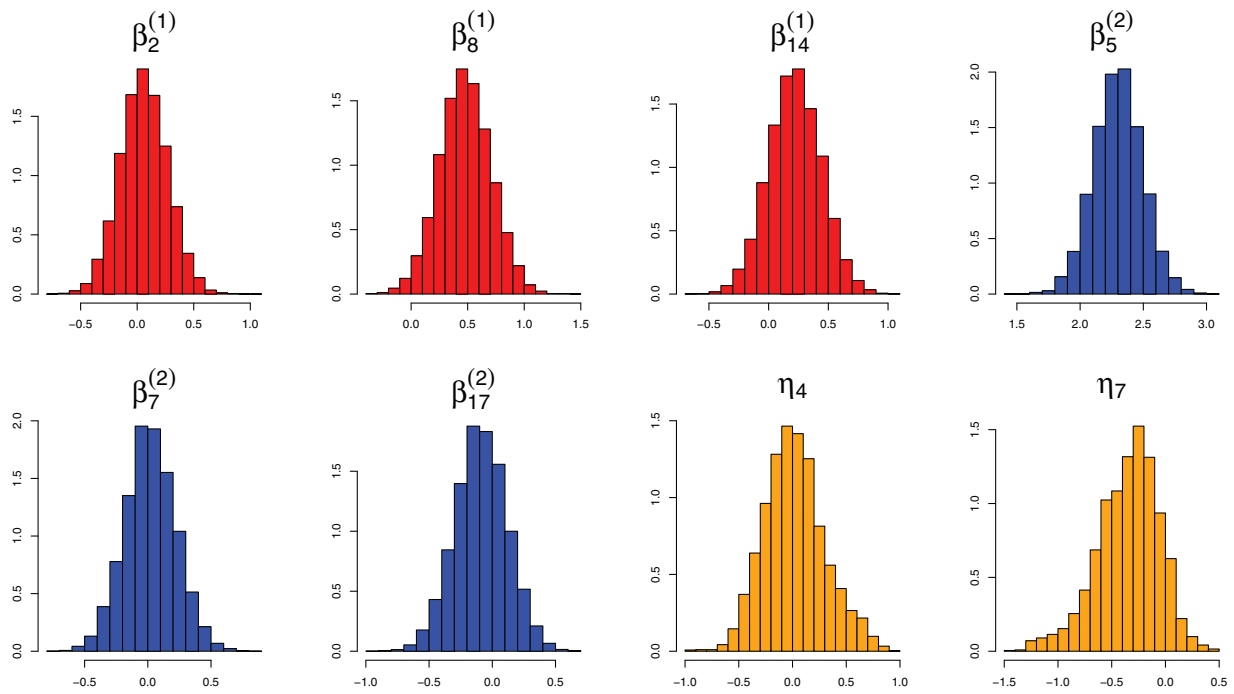


Figure 5. Histograms for the selected parameters of one replicate from Setup 2 when  $p = 20$  and  $s = 0.2$ .

on explaining only the results from Table 1. Following are our observations from the simulation.

- BHQQ(1) and BHQQ(2) have very similar performances, indicating that the proposed Bayesian hierarchical model is not sensitive to the two options  $(v, \delta^2, a, b)$ .
- In terms of RMSPE for Setup 1 the BHQQ is inferior to SM(F) and SM(B), but this is expected. On the one hand,  $\beta^{(1)}$  and  $\beta^{(2)}$  are similar, which implies the quantitative response  $Y$  does not depend on the qualitative response  $Z$ , so the separate models should be sufficient. On the other hand, the BHQQ involves twice as many unknown parameters for  $Y$  as the separate models, which could lead to the inferior parameter estimates. The good news is that the RMSPE of BHQQ outperforms the separate models for Setups 2 – 4, in which  $\beta^{(1)}$  is different from  $\beta^{(2)}$ , reflecting the dependency of  $Y$  on  $Z$ . The proposed model is especially significantly superior to the separate models for Setups 2 and 4. One possible explanation is that when the significant variables of  $\beta^{(1)}$  and  $\beta^{(2)}$  are the same (i.e., Setups 1 and 3), the separate models are still able to distinguish between the significant and insignificant variables, despite the fact that  $\beta_1$  and  $\beta_2$  are different (Setup 3). However, when the significant variables are different (i.e., Setups 2 and 4), there is a stronger dependency of  $Y$  on  $Z$ ; thus the separate models

have much more difficulty in identifying the significant variables.

- We also note that the BHQQ is comparable to, and sometimes slightly better than, the SM(F) and SM(B) with respect to ME. This is because both BHQQ and separate models fit the binary response  $Z$  based on the marginal distribution  $p(\mathbf{z}|\boldsymbol{\eta})$ .
- In terms of  $\gamma_{lm}$ , the proposed Bayesian hierarchical model remarkably dominates SM(F). It is also better than SM(B) for Setups 1 and 3 in the setting  $s = 0.2$ , and it is comparable in the setting  $s = 0.5$ . For Setups 2 and 4 the BHQQ significantly outperforms SM(B) for the same reason as mentioned above—the separate models do not perform well because of the dependency between  $Y$  and  $Z$ .
- For the parameter estimation accuracy  $L_2(\beta^{(1)})$  and  $L_2(\beta^{(2)})$  the BHQQ is inferior to the SM(F) and SM(B) for Setup 1 but greatly outperforms the separate methods for Setups 2 – 4. This is due to the same reason as the RMSPE for Setup 1.
- For  $\gamma_{logit}$  the proposed model is similar to SM(B) for all settings. Compared with SM(F) the proposed model gives superior performance in the setting  $s = 0.2$  and is similar to SM(F) in the setting  $s = 0.5$ .
- In addition, we also notice that the proposed model is slightly worse than the separate models with respect to  $L_2(\boldsymbol{\eta})$ . One possible reason could be the generalized linear model plus LASSO slightly outperforms the Bayesian version of the generalized linear model.



**Table 3.** Measured predictor variables and quality responses in the lapping experiment.

Type	Name	Definition
Lapping Process Variables	$x_1$ : Pressure ( $N/m^2$ )	The high pressure of the upper to lower plate
	$x_2$ : Rotation (Rpm)	The rotation speed
	$x_3$ : LPTime (Sec.)	The time for low pressure
	$x_4$ : HPTime (Sec.)	The time for high pressure
Quality Covariates before Lapping	$x_5$ : CTHK0 ( $\mu m$ )	Central thickness of wafers
	$x_6$ : TTVO ( $\mu m$ )	Total thickness variation of wafers
	$x_7$ : TIRO ( $\mu m$ )	Total indicator reading of wafers
	$x_8$ : STIRO ( $\mu m$ )	Site total indicator reading (STIR) of wafers
	$x_9$ : BOWO ( $\mu m$ )	Deviation of local warp at the center of wafers
	$x_{10}$ : WARPO ( $\mu m$ )	Maximum of local warp of wafers
QQ Responses	$y$ : TTV ( $\mu m$ )	Continuous total thickness variation of wafers
	$z$ : STIR indicator (0 or 1)	Binary indicator for the conformity of STIR

To check the convergence property of MCMC chains for our interests  $\beta^{(1)}$ ,  $\beta^{(2)}$ , and  $\eta$  we use the Gelman-Rubin diagnostic through the coda package. Here we show only the multivariate scale reduction factor from one randomly selected simulation out of the 100 from Setup 2, where  $p=20$  and  $s=0.2$ . Their values are 1.004, 1.003, and 1.209 for  $\beta^{(1)}$ ,  $\beta^{(2)}$ , and  $\eta$ , respectively. In addition, Figure 5 displays the histograms for the posterior samples of some parameters after the burn-in period.

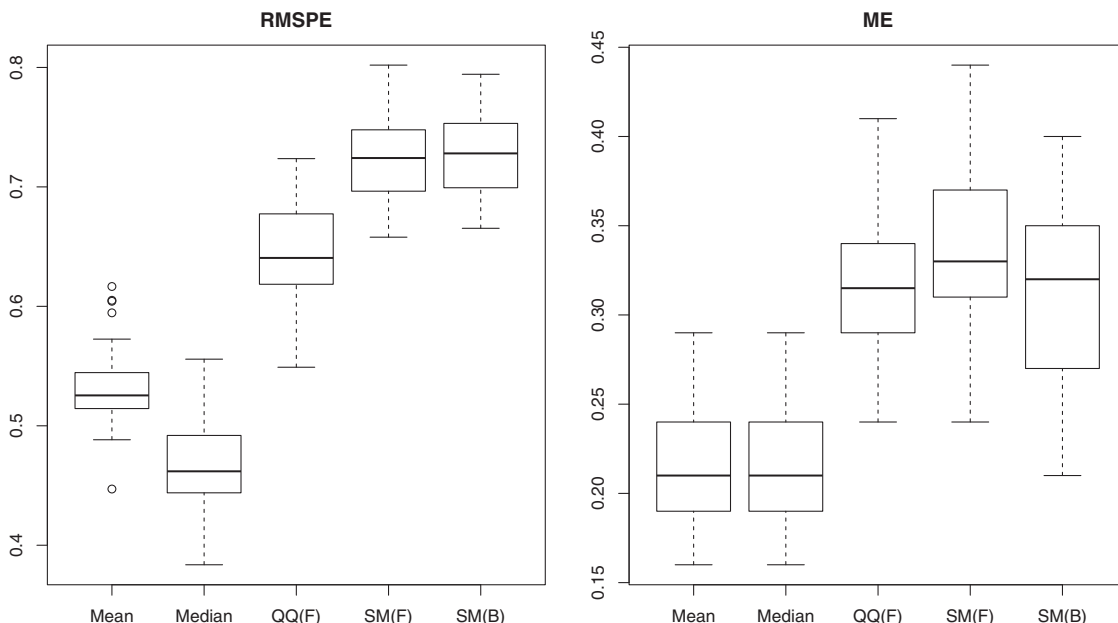
**5. Case study: Lapping experiment**

We study the experiment of the lapping process described in Section 1. The experimental data include  $n = 450$  observations,  $p = 10$  predictor variables  $x = (x_1, \dots, x_p)'$ , and two responses as described in Table 3. The quantitative response is the TTV, and the binary response is the indicator variable denoting

whether or not the STIR is larger than the tolerance. The basis function  $f(x)$  here is just the intercept term and predictor variables  $x_1, \dots, x_p$ , (i.e.,  $f(x) = (1, x_1, \dots, x_p)'$ ).

We set the number of MCMC iterations  $B = 10,000$  with burn-in period  $b = 1,000$ . The initial values of the hyperparameters  $(r_{1,0}, r_{2,0}, r_{3,0}, \tau_{1,0}^2, \tau_{2,0}^2)$  are set as  $(0.3, 0.3, 0.3, 1.5, 3)$ . We set the model tuning parameter  $(\nu, \delta^2, a, b) = (2, 2, 0.5, 0.5)$ . To compare the performance of the BHQQ with SM(F), SM(B), and the one proposed by Deng and Jin (2015) (denoted by QQ(F)), we randomly split the whole data set into two groups: a training set of 350 samples and a testing set of 100 samples. For each random split we fit the four methods and make predictions. We repeat this 50 times.

Figure 6 shows the boxplots of the RMSPE values for  $Y$  and the ME values for  $Z$ . “Mean” or “Median” indicate whether we use the mean or median of the



**Figure 6.** Box plots of two prediction measures for lapping data for different methods.

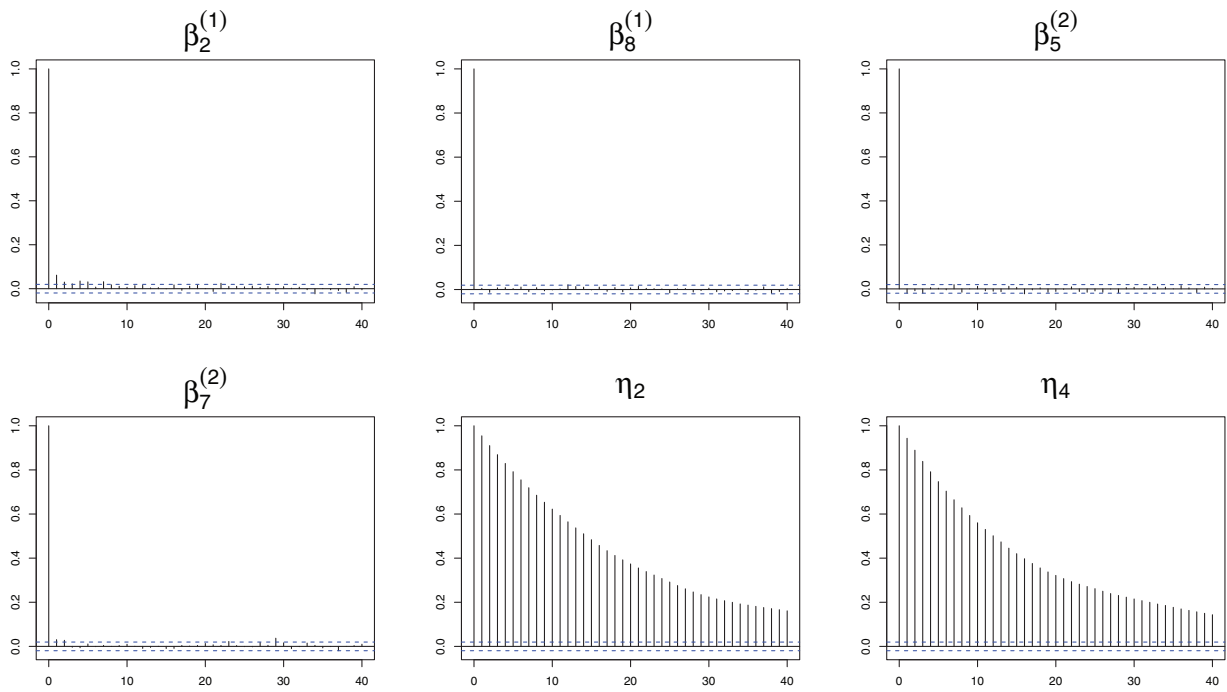


Figure 7. ACF plots for the selected parameters for the lapping data.

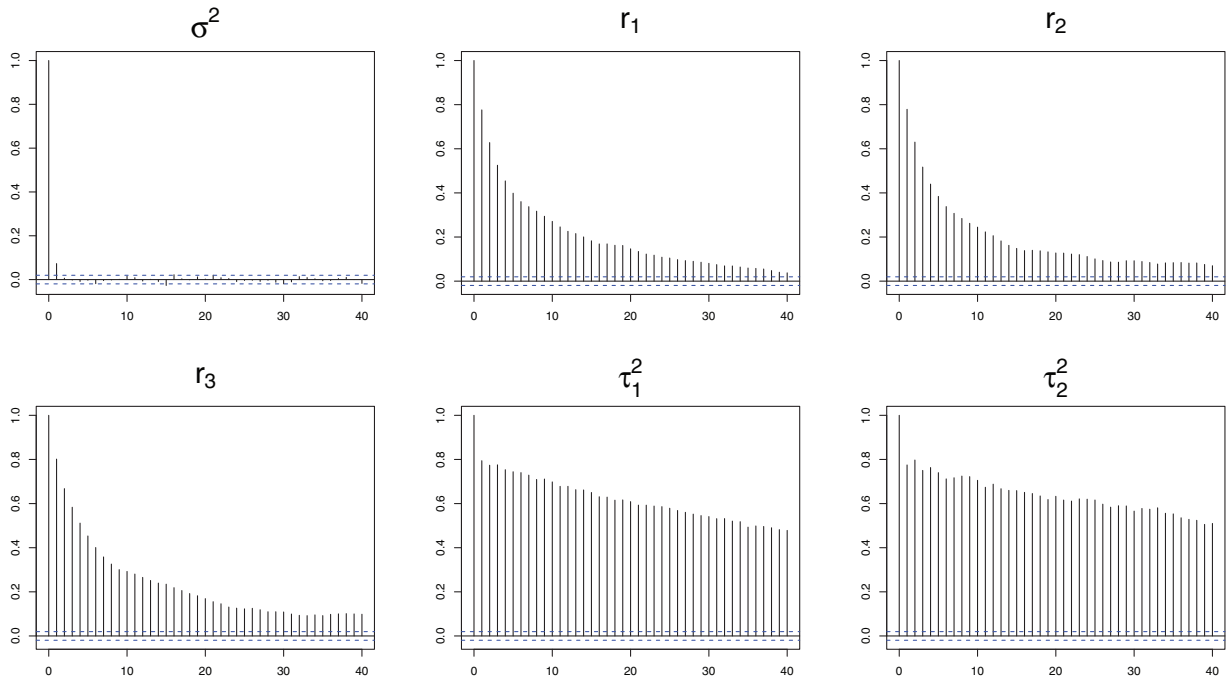
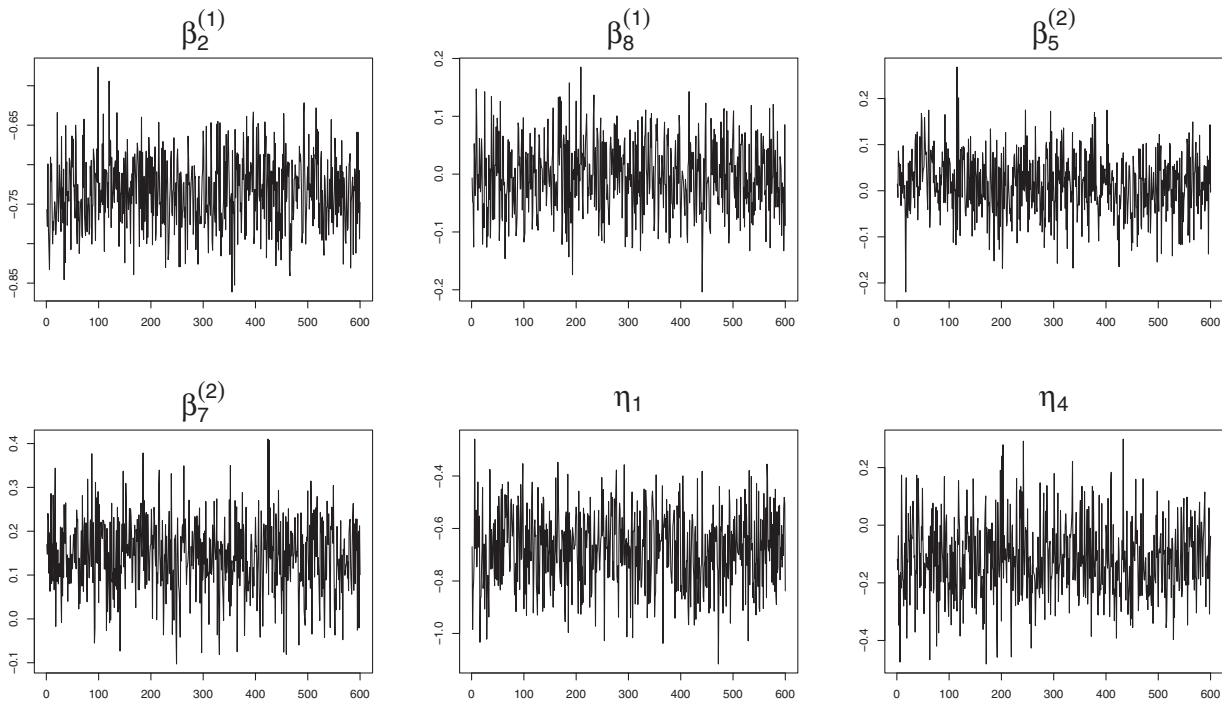


Figure 8. ACF plots for  $(\sigma^2, r_1, r_2, r_3, \tau_1^2, \tau_2^2)$ .

9,000 posterior predictive samples as the predicted values for  $Y$  and  $Z$ . We compare them with the RMSPE and ME of three alternative methods. For this experiment the sample size is greatly larger than the model size; thus all the frequentist approaches can be used. According to Figure 6, SM(F) and SM(B) perform comparably in terms of RMSPE and ME. QQ(F) has a lower prediction error than SM(F) and

SM(B) since it considers the dependency of  $Y$  on  $Z$ . Nevertheless, BHQQ greatly outperforms QQ(F) in terms of both RMSPE and ME. Overall, the proposed Bayesian hierarchical model shows obvious improvement in prediction accuracy. Note that using the median of the posterior predictive samples is more accurate than the mean for the  $Y$  response.



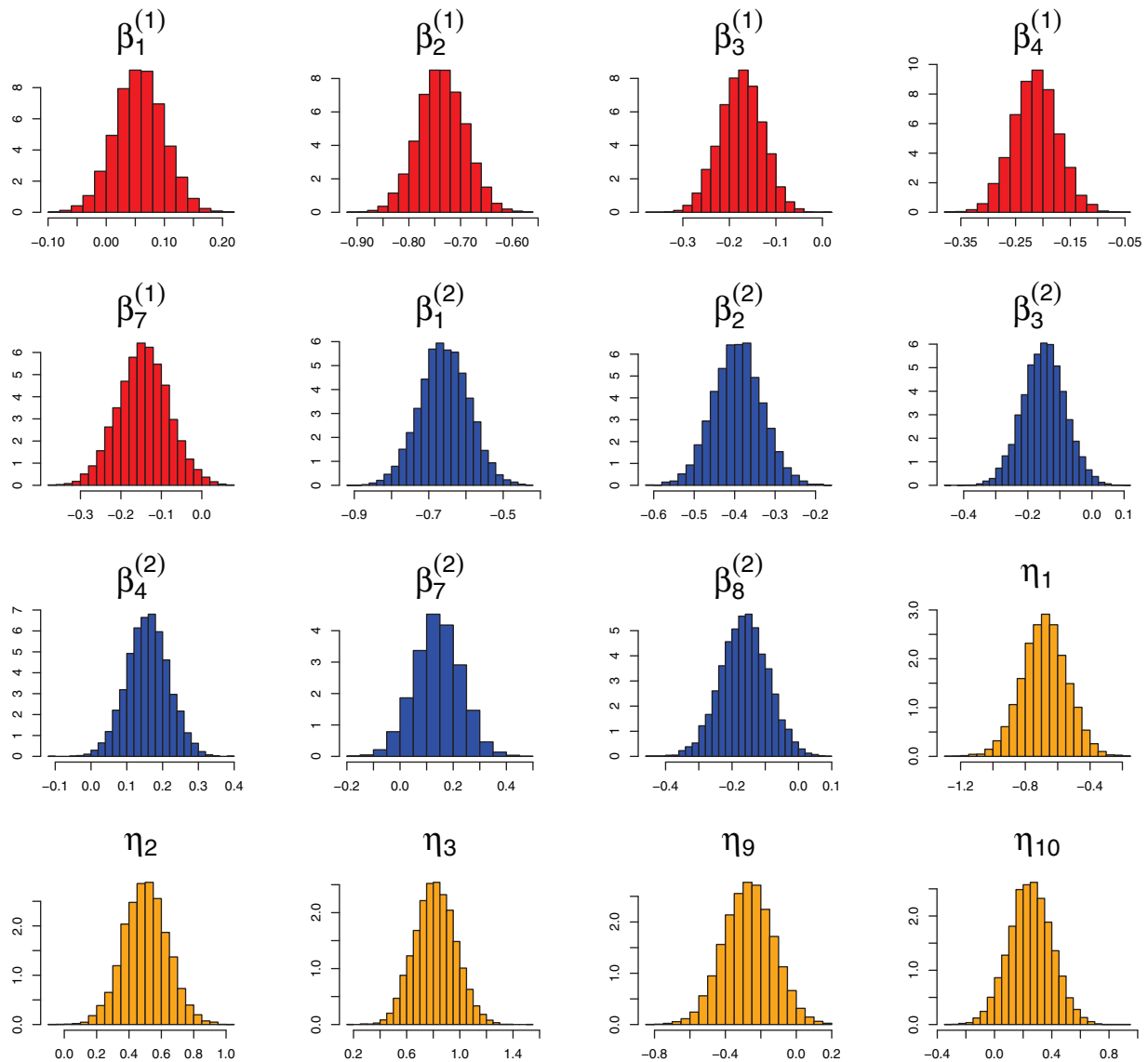
**Figure 9.** Trace plots for the selected parameters for the lapping data.

It is worth pointing out that the MEs of all four methods are relatively large, even though the proposed BHQQ method has the smallest ME. A possible explanation is that the binary response of the lapping experiment is quite noisy, which is confirmed by the QQ(F), SM(F), and SM(B) methods and by Deng and Jin (2015). Compared with SM(F) and SM(B), the BHQQ method has reduced the misclassification error, but such improvement is not as significant as it is for the continuous response. This is due to the conditional structure  $Y|Z$ . Additional information from the responses  $z$  is included in the model of  $Y$ , whereas  $Z$  is still modeled marginally.

Next, we look into the analysis results of the BHQQ model for only one randomly split training and testing data set. There are 34 observations with  $Z = 0$  and 66 with  $Z = 1$  in this particular testing set. For clear illustration we reorder the testing data such that the first 34 observations correspond to  $Z = 0$ . Figures 7 and 8 are the autocorrelation function (ACF) plots for some selected parameters and  $(\sigma^2, r_1, r_2, r_3, \tau_1^2, \tau_2^2)$ . The quick drop off of ACF in these plots, especially of  $\beta^{(1)}$ ,  $\beta^{(2)}$ , and  $\sigma^2$ , implies the fast convergence of the Gibbs sampling iteration. Figure 9 presents the trace plots for some selected parameters. After the burn-in period the traces of the parameters fluctuate around the means with relatively stable variation, which further confirms that the parameter posterior sampling converges. The rest of the parameters have similar patterns; hence their plots are omitted.

Figure 10 shows the histograms of the MCMC samples (after the burn-in ones have been removed) corresponding to the significant variables. Figure 11 displays the box plots of the posterior samples of each element of  $\beta^{(1)}$ ,  $\beta^{(2)}$ , and  $\eta$  from which we can easily identify the significant ones. Also, it is clear that  $\beta^{(1)}$  and  $\beta^{(2)}$  have very different distributions, implying that  $Y$  strongly depends on  $Z$ ; thus the proposed Bayesian hierarchical model is able to account for this relationship.

We plot the true observations  $y$  against the predicted values  $\hat{y}$  of the testing set in Figure 12. The prediction is relatively accurate since most points are around the line of  $y = \hat{y}$ . Figure 13 shows the true observations  $y$  (black dashed line) in the testing set and their 95 percent prediction intervals (red dotted lines) from the posterior predictive samples. Among them, 61 out of the 100 observations are correctly covered by the intervals. Figure 14 plots the estimated probability (black dashed line) and its 95 percent prediction intervals (red dotted lines) for testing the data set. The first 34 binary responses are equal to zero. They are on the left of the black solid line of index = 34. The horizontal solid line represents  $p = 0.5$ . If the estimated probability is above the line  $p = 0.5$ , its predicted value is set to be one and zero otherwise. We are more confident assigning values one or zero when the prediction intervals are totally above or below the line  $p = 0.5$ . These predicted values are shown in Figure 15—the red triangles are the misclassifications.



**Figure 10.** Histograms for the parameter estimates of the significant variables for the lapping data. Red color presents elements of  $\beta^{(1)}$ , blue for  $\beta^{(2)}$ , and yellow for  $\eta$ .

The proposed model incorrectly classifies 12 good wafers (STIR indicator is zero) as bad (STIR indicator is one), and five bad wafers as good. The misclassification error rate is 17 percent. In contrast, QQ(F) incorrectly gives 21 good wafers as bad and wrongly predicts 9 bad wafers as good. The misclassification rate of QQ(F) is 30 percent, which is worse than the proposed model.

### 6. Conclusion

We propose a Bayesian hierarchical model to analyze the experiments whose outputs contain a continuous response and a binary response. The proposed method can jointly model the two types of responses and discover the dependency between them, which exceeds

the simple separate modeling of the two types of responses. The latest existing method developed by Deng and Jin (2015) can jointly model the two responses via maximizing the constrained likelihood, but it has two restrictions that are discussed in the Introduction. The proposed Bayesian framework can easily overcome these two restrictions. In addition, its estimation, prediction, and inference accuracy are comparable to or much better than all the existing methods, including Deng and Jin (2015), as shown in Section 4.

In this article we focus only on the simplest case of the quantitative and qualitative system (i.e., a single normal distributed quantitative response and a single binary qualitative response). The proposed Bayesian hierarchical modeling approach can be extended to

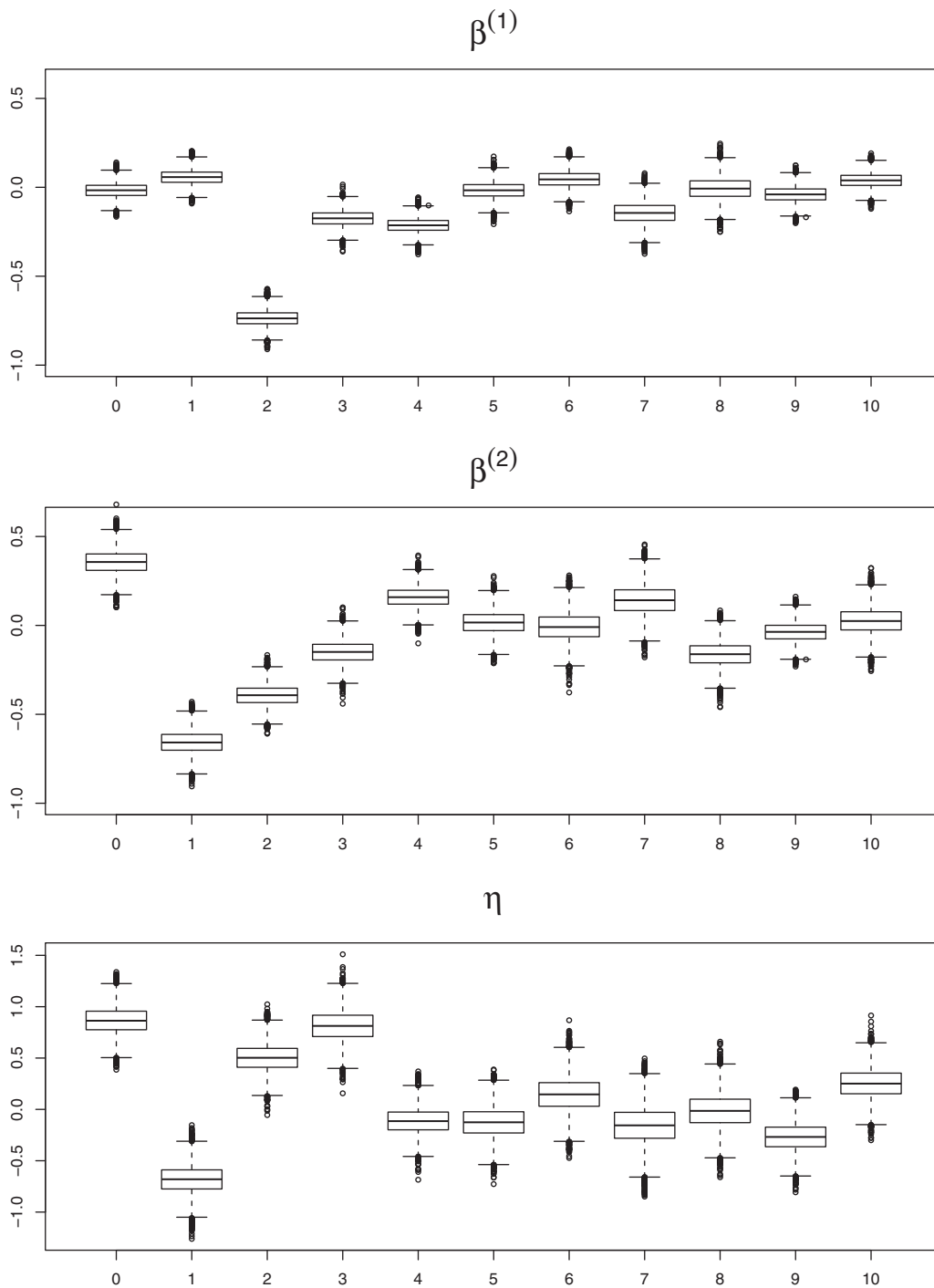


Figure 11. Box plots for the parameter estimates for the lapping data.

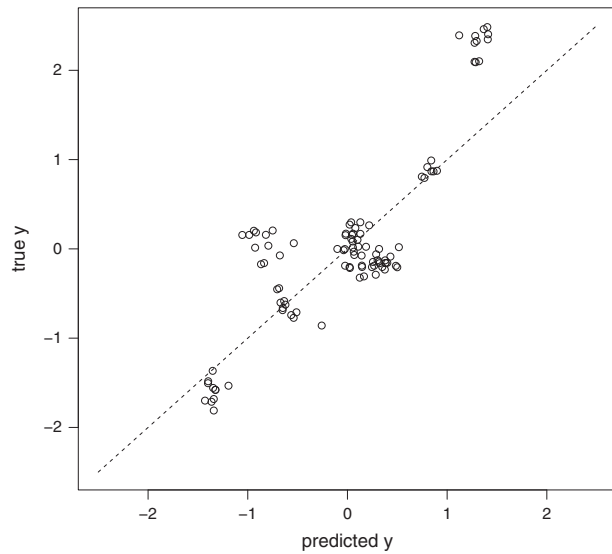
some more complicated situations. For example, when there are multivariate quantitative responses we assume a multivariate normal distribution model accordingly and assume the correlation between the different quantitative responses if needed. When the qualitative response has finite but more than two levels we can assume the multinomial logistic regression model instead of the Bernoulli logistic regression model. Some more complicated cases may occur, such

as multiple qualitative responses, in which the proposed approach cannot be easily extended. We plan to pursue these topics in our future research.

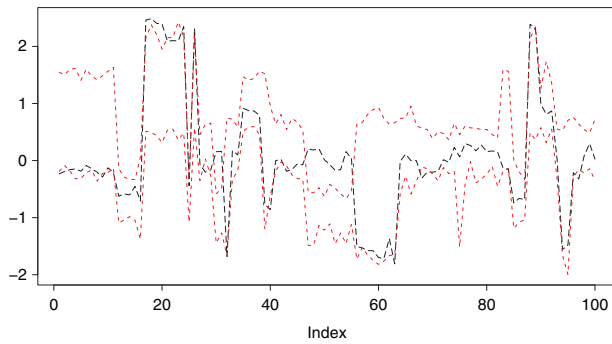
### About the authors

**Dr. Lulu Kang** is an Associate Professor of the Department of Applied Math at Illinois Institute of Technology (IIT). She obtained an M.S. in Operations

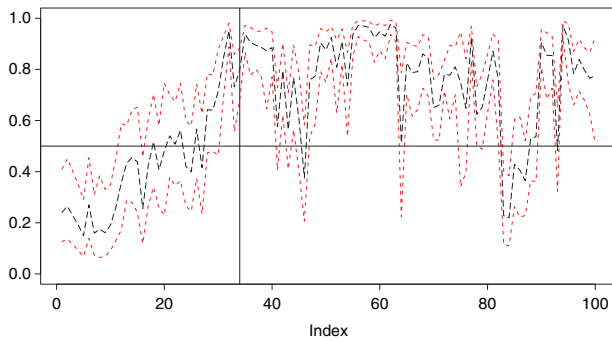




**Figure 12.** Plot of the true observations  $y$  against the predicted  $\hat{y}$  of the testing data set.

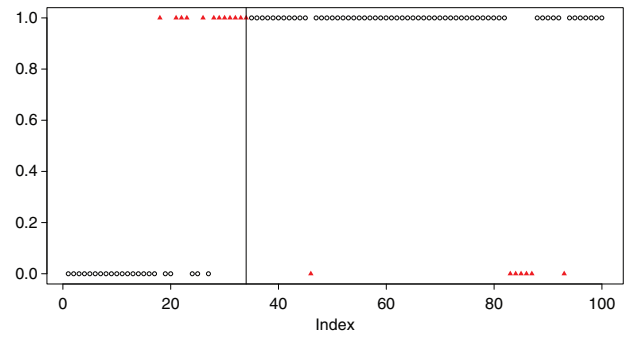


**Figure 13.** Plot of the true observations  $y$  of testing data set and the corresponding 95 percent prediction intervals.



**Figure 14.** Plot of the estimated probabilities and the corresponding 95 percent prediction intervals using the testing data set.

Research and a Ph.D. in Industrial Engineering from the Stewart School of Industrial and Systems Engineering at Georgia Tech. Dr. Kang has worked on various areas in Statistics, including uncertainty quantification, statistical design and analysis of experiments, Bayesian computational statistics, etc. She has



**Figure 15.** Plot of the estimated qualitative responses  $Z$  of testing data set. The black circles represent the correct classifications; the red triangles indicate the misclassifications.

collaborated with researchers in mechanical, manufacturing, environmental engineering fields. Dr. Kang is also the co-founder and Associate Program Director of the Data Science Program at IIT.

**Dr. Xiaoning Kang** holds a Ph.D. in statistics from Virginia Tech. He is now an Assistant Professor at International Business College in Dongbei University of Finance and Economics, China. His research interests include high dimensional data analysis, large covariance matrix estimation, and statistical methodology with financial applications.

**Dr. Xinwei Deng** is an Associate Professor in the Department of Statistics at Virginia Tech. He received his Ph.D. from Georgia Institute of Technology in 2009. Dr. Deng’s research interests includes data analytics, machine learning, design of experiments, and interface between experimental design and machine learning. Dr. Deng actively collaborates with researchers in engineering, nanotechnology, bioinformatics, and environmental sciences. He has over 35 peer-reviewed top journal and conference publications. Dr. Deng is currently an associate editor in *Technometrics*, *International Statistics Review*, and *IISE Transactions*. Dr. Deng’s research has been funded by the National Science Foundation and industrial collaborations.

**Dr. Ran Jin** is an Assistant Professor and the Director of Laboratory of Data Science and Visualization at the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received his Ph.D. degree in Industrial Engineering from Georgia Tech, Atlanta, his Master’s degrees in Industrial Engineering, and in Statistics, both from the University of Michigan, Ann Arbor, and his Bachelor’s degree in Electronic Engineering from Tsinghua University, Beijing. His research focuses on Data Fusion in Smart Manufacturing, including the integration of different-types of data sets (e.g., ensemble models), variables (e.g., quantitative and qualitative models), and information (e.g., product quality and equipment reliability)

for synergistically modeling, monitoring and control of manufacturing processes and systems. He is currently serving as an Associate Editor for *IIE Transactions, Focus Issue on Design and Manufacturing*.

## Funding

This research was supported by U.S. National Science Foundation Grants CMMI-1435902 and CMMI-1435996.

## References

- Ai, M., Kang, L., and V. R. Joseph. 2009. Bayesian optimal blocking of factorial designs. *Journal of Statistical Planning and Inference* 139 (9):3319–3328.
- Bello, N. M., J. P. Steibel, and R. J. Tempelman. 2012. Hierarchical Bayesian modeling of heterogeneous cluster- and subject-level associations between continuous and binary outcomes in dairy production. *Biometrical Journal* 54 (2):230–248.
- Breiman, L. 1995. Better subset regression using the non-negative garrote. *Technometrics* 37 (4):373–384.
- Casella, G., and E. I. George. 1992. Explaining the Gibbs sampler. *The American Statistician* 46 (3):167–174.
- Catalano, P. J., and L. M. Ryan. 1992. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* 87 (419):651–658.
- Chen, M. H., and J. G. Ibrahim. 2003. Conjugate priors for generalized linear models. *Statistica Sinica* 13 (2):461–476.
- Chen, S., D. M. Witten, and A. Shojaie. 2014. Selection and estimation for mixed graphical models. *Biometrika* 102 (1):47–64.
- Cox, D. R., and N. Wermuth. 1992. Response models for mixed binary and quantitative variables. *Biometrika* 79 (3):441–461.
- Deng, X., and R. Jin. 2015. QQ models: Joint modeling for quantitative and qualitative quality responses in manufacturing systems. *Technometrics* 57 (3):320–331.
- Dunson, D. B., and A. H. Herring. 2005. Bayesian latent variable models for mixed discrete outcomes. *Biostatistics* 6 (1):11–25.
- Fitzmaurice, G. M., and N. M. Laird. 1995. Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association* 90 (431):845–852.
- Fitzmaurice, G. M., and N. M. Laird. 1997. Regression models for mixed discrete and continuous responses with potentially missing values. *Biometrics* 53 (1):110–122.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian data analysis*. 3rd ed., Vol. 2. Boca Raton, FL: CRC Press.
- Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7 (4):457–472.
- Guglielmi, A., F. Ieva, A. M. Paganoni, and F. A. Quintana. 2016. A semiparametric Bayesian joint model for multiple mixed-type outcomes: An application to acute myocardial infarction. *Advances in Data Analysis and Classification* 12 (2):399–423.
- Holmes, C. C., and L. Held. 2006. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* 1 (1):145–168.
- Joseph, V. R., 2006. A Bayesian approach to the design and analysis of fractionated experiments. *Technometrics* 48 (2):219–229.
- Kang, L., X. Deng, and R. Jin. 2015. Bayesian d-optimal design of experiments with continuous and binary responses. Revision invited for *Journal of American Statistical Association*.
- Kang, L., and V. R. Joseph. 2009. Bayesian optimal single arrays for robust parameter design. *Technometrics* 51 (3):250–261.
- Ning, Y. D., Y. Z. Bian, and B. Liu. 2012. Improving a lapping process using robust parameter design and run-to-run control. *Journal of the Chinese Institute of Industrial Engineers* 29 (2):111–124.
- Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R News* 6 (1):7–11.
- Roberts, G. O., and A. F. Smith. 1994. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and Their Applications* 49 (2):207–216.
- Sammel, M. D., L. M. Ryan, and J. M. Legler. 1997. Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59 (3):667–678.
- Weiss, R. E., J. Jia, and M. A. Suchard. 2011. A Bayesian model for the common effects of multiple predictors on mixed outcomes. *Interface Focus* 1 (6):886–894.
- Wu, C. J., and M. S. Hamada. 2011. *Experiments: Planning, analysis, and optimization*. Vol. 552. Hoboken, NJ: John Wiley & Sons.
- Yang, E., Y. Baker, P. Ravikumar, G. Allen, and Z. Liu. 2014. Mixed graphical models via exponential families. Paper presented at the Seventeenth International Conference on Artificial Intelligence and Statistics, Reykjavik, Iceland, JMLR: W&CP vol. 33. 1042–1050.
- Zhang, L., K. Wang, and N. Chen. 2015. Monitoring wafers' geometric quality using an additive Gaussian process model. *IIE Transactions* 48 (1):1–15.
- Zhao, H., R. Jin, S. Wu, and J. Shi. 2011. PDE-constrained Gaussian process model on material removal rate of wire saw slicing process. *Journal of Manufacturing Science and Engineering* 133 (2):021012.