Technometrics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

Bayesian Sparse Regression for Mixed Multi-Responses with Application to Runtime Metrics Prediction in Fog Manufacturing

Xiaoyu Chen, Xiaoning Kang, Ran Jin & Xinwei Deng

To cite this article: Xiaoyu Chen, Xiaoning Kang, Ran Jin & Xinwei Deng (2022): Bayesian Sparse Regression for Mixed Multi-Responses with Application to Runtime Metrics Prediction in Fog Manufacturing, Technometrics, DOI: 10.1080/00401706.2022.2134928

To link to this article: https://doi.org/10.1080/00401706.2022.2134928

TECHNOMETRICS
2022, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/00401706.2022.2134928

Taylor & Francis
Taylor & Francis Group

Bayesian Sparse Regression for Mixed Multi-Responses with Application to Runtime Metrics Prediction in Fog Manufacturing

Xiaoyu Chen[a], Xiaoning Kang[b], Ran Jin[c], and Xinwei Deng[d]

[a]Department of Industrial Engineering, University of Louisville, Louisville, KY; [b]International Business College and Institute of Supply Chain Analytics, Dongbei University of Finance and Economics, Dalian, China; [c]Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA; [d]Department of Statistics, Virginia Tech, Blacksburg, VA

ABSTRACT

Fog manufacturing can greatly enhance traditional manufacturing systems through distributed Fog computation units, which are governed by predictive computational workload offloading methods under different Industrial Internet architectures. It is known that the predictive offloading methods highly depend on accurate prediction and uncertainty quantification of runtime performance metrics, containing multivariate mixed-type responses (i.e., continuous, counting, binary). In this work, we propose a Bayesian sparse regression for multivariate mixed responses to enhance the prediction of runtime performance metrics and to enable the statistical inferences. The proposed method considers both group and individual variable selection to jointly model the mixed types of runtime performance metrics. The conditional dependency among multiple responses is described by a graphical model using the precision matrix, where a spike-and-slab prior is used to enable the sparse estimation of the graph. The proposed method not only achieves accurate prediction, but also makes the predictive model more interpretable with statistical inferences on model parameters and prediction in the Fog manufacturing. A simulation study and a real case example in a Fog manufacturing are conducted to demonstrate the merits of the proposed model.

ARTICLE HISTORY
Received July 2021
Accepted September 2022

KEYWORDS
Graphical model; Mixed responses; Spike-and-slab prior; Variable selection

1. Introduction

Fog computing (also referred to as Edge computing) techniques have served as an important role in Industrial Internet of things (IIoT) for smart manufacturing systems. It provides local and distributed computation capabilities. The concept of Fog manufacturing is defined on integrating a Fog computing network with interconnected manufacturing processes, facilitates, and systems. With local computation units (i.e., Fog units) close to the manufacturing processes, the Cloud-based centralized computation architecture can be evolved to a Cloud-Fog collaborative computation to provide higher responsiveness and significantly lower time latency (Wu et al. 2017; Zhang et al. 2019). There is a tradeoff between the local computing efficiency on a Fog unit and the global collaborative efficiency of the centralized Cloud. Specifically, the speciality of Fog units can significantly speedup the local computations, but it can pose significant challenges for the Cloud to assign the computation tasks and supervise the heterogeneous Fog units. Besides, fluctuated computation capability of the Fog units and intermittent communication conditions among the Fog units and the Cloud make it even harder for the collaboration (Zhang, Niyato, and Wang 2015). Therefore, computation offloading methods have been widely investigated to enable efficient collaboration between the Fog units and the Cloud with the consideration of constraints on resources.

In Fog manufacturing, the runtime performance metrics are often multivariate with mixed types (Chen et al. 2018). These metrics include the CPU utilization (i.e., continuous response), temperature of the CPU (i.e., continuous response), the number of computation tasks executed within a certain time period (i.e., counting response), and whether the memory utilization exceeds certain thresholds (i.e., binary response). Prediction and uncertainty quantification of these metrics are essential to support the computation in the Fog manufacturing, advancing analytics and optimization for high responsiveness and reliability (Wu et al. 2017; Zhang et al. 2019). Based on the runtime performance metrics of these Fog nodes, the Fog computing can dynamically assign computation tasks to different Fog nodes (Chen et al. 2018). The manufacturing must provide responsive and reliable computation services by meeting all requirements in runtime performance metrics. It is thus of great importance to accurately predict runtime performance metrics of Fog nodes and quantify the uncertainty of prediction in task assignment and offloading problems.

As the runtime performance metrics are multivariate with mixed types, a simple method is to model each individual metric separately. Clearly, such an approach overlooks the dependency relationship among the metrics, resulting in inaccurate prediction associated with high uncertainty. For example, as the increment in the executed number of computation tasks

CONTACT Xinwei Deng *xdeng@vt.edu* Department of Statistics, Virginia Tech, Blacksburg, VA 24061.
Xiaoyu Chen and Xiaoning Kang have contributed equally to this work.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/TECH*.

per minute (i.e., counting response), the CPU utilization and temperature (i.e., continuous responses) will increase. Quantifying such dependency among mixed responses is expected to improve the prediction accuracy. Moreover, by only providing point estimation of mixed responses, the model prediction may not be trustworthy for those with high prediction variance. Therefore, it calls for a joint model for the mixed responses with uncertainty quantification. Toward predictive offloading, the objective is to jointly fit the mixed runtime performance metrics with the capability of statistical inferences to quantify uncertainties of the predicted metrics in Fog manufacturing.

In this work, we propose a Bayesian sparse multivariate regression for mixed responses (BS-MRMR) to achieve accurate model prediction and, more importantly, to obtain proper statistical inferences of the responses. The use of Bayesian estimation naturally enables uncertainty quantification of model prediction. Both group sparsity and individual sparsity are imposed on regression coefficients via proper spike-and-slab priors. The group structures often occur in the runtime performance metrics prediction problem when the metrics at the next time instance are regressed on two groups of predictors: the features extracted from the current and previous metrics (i.e., Group 1) and the covariates of the computation tasks (i.e., Group 2). On the other hand, not all predictors are important within each group. Hence, the individual sparsity is also induced for better estimation of model coefficients. Moreover, the proposed method considers the conditional dependency among multiple responses by a graphical model using the precision matrix, where a spike-and-slab prior is used to enable the sparse estimation of the graph. A Gibbs sampling scheme is then developed to efficiently conduct model estimation and inferences for the proposed BS-MRMR method. The proposed BS-MRMR model not only achieves accurate prediction, but also makes the predictive model more interpretable in the Fog manufacturing. Note that one can consider a two-step Bayesian method to model the multivariate mixed responses Bradley (2022), where the first step transforms the multivariate mixed-responses to continuous responses, and the second step models the transformed responses. However, the obtained model coefficients are less interpretable since the transformation typically change the scale of the original responses.

Different from the recent work of Kang et al. (2021) on a penalized regression for multivariate mixed responses, the proposed BS-MRMR is a Bayesian approach with the following key novelty. First, Kang et al. (2021) only imposes individual sparsity while the BS-MRMR model takes into account of both the group and individual sparsity. Second, the model introduced by Kang et al. (2021) cannot provide statistical inferences, such as prediction intervals for the responses due to their complicated parameter estimation procedure. In contrast, the proposed BS-MRMR model is able to quantify the uncertainty of the estimated parameters and predicted responses within the Bayesian framework. It provides a comprehensive information of prediction and uncertainty quantification to support the predictive offloading in Fog manufacturing. Third, a careful investigation of the posterior distribution makes the computation of the Gibbs sampling efficient for model estimation and inference.

The remainder of this work is organized as follows. The proposed BS-MRMR model and the Gibbs sampling scheme are detailed in Section 3. A simulation study is conducted to validate the BS-MRMR model in Section 4. Section 5 describes a real case study in Fog manufacturing. Section 6 concludes this work with some discussions of future directions.

2. Literature Review

The joint modeling of mixed responses has attracted great attention in the literature. Various existing studies focused on the bivariate responses. For example, Fitzmaurice and Laird (1995) considered a bivariate linear regression model with a continuous and a binary response via joint likelihood estimation. Yang et al. (2007) proposed to jointly fit a continuous and a counting response, and evaluated the correlation between the bivariate responses varying over time through a likelihood ratio test. These methods usually factorize the joint distribution of two responses as the product of a marginal and a conditional distribution (Cox and Wermuth 1992), which cannot be easily generalized for multivariate mixed responses in real applications such as Fog manufacturing. Another direction of handling the continuous and discrete variables is to consider the underlying latent variables for the discrete responses, and then assume a multivariate normal distribution for such latent variables together with other continuous responses. For example, Regan and Catalano (1999) introduced a latent variable with a probit link function for a binary response and jointly modeled the continuous response and the latent variable via a bivariate normal distribution. More related works include McCulloch (2008), Deng and Jin (2015), Wu, Deng, and Ramakrishnan (2018), and Kang et al. (2021). The advantage of introducing latent variables to characterize discrete responses lies mainly in the well-defined correlation measures among multivariate normal responses. Therefore, the hidden association between mixed responses can be quantified by this correlation. However, such models involving latent variables are often computationally expensive, especially when the number of predictor variables is large.

Modeling the mixed responses under the Bayesian framework is also studied in the literature. For example, Fahrmeir and Raach (2007) fitted ordinal and normal responses via a Bayesian latent variable method, where covariate effects on the latent variables were modeled through a semiparametric Gaussian regression model. Yeung et al. (2015) studied a dose-escalation procedure in clinical trials by a Bayesian approach, where a logistic regression and a linear log-log relationship were used, respectively, to model the binary and continuous responses. Kang et al. (2018) proposed to fit the binary response conditioned on the continuous response, where proper priors were used for enhancing model interpretation. However, few Bayesian works have been conducted for the multivariate mixed responses. Li, Pan, and Belcher (2016) introduced a Bayesian conditional joint random-effects model for fitting longitudinal data with normal, binary and ordinal responses by using latent variables for each response. More Bayesian methods can be found in Dunson (2000), Zhou et al. (2006), Stamey, Natanegara, and Seaman Jr (2013), Hwang and Pennell (2014), DeYoreo and Kottas (2018), among others.

There are few works of theoretical investigation on models of mixed responses under the framework of either conditional

models or latent variables. In the simple case where there are only one binary and one continuous responses, Kürüm et al. (2016) adopted latent variable to characterize the binary response and studied the asymptotic normality of their estimator. Recently, Kang et al. (2022) developed a generative model framework in which the continuous response is fitted based on the multivariate normal property, and the multi-class response is modeled and predicted via the linear discriminant analysis. They established the asymptotic properties of their estimator in terms of both the classification accuracy for the multi-class response and the prediction accuracy for the continuous response under some regularity conditions.

In addition, the proper regularization on model parameters is often used in the joint modeling of mixed responses for high-dimensional data to improve the model interpretation. Kang et al. (2021) proposed to fit data with multiple mixed responses by imposing L_1 penalties on the negative log-likelihood function and conducted the parameter estimation based on the EM algorithm. Under the Bayesian framework, the spike-and-slab prior is commonly used for inducing the sparsity in regression models (Wagner and Tüchler 2010). In our application of runtime performance metrics prediction problem, particularly, the group variable selection is necessary since predictor variables are naturally grouped by different components (e.g., CPU, RAM, etc.) of Fog units. Hence, the model coefficients may not be properly estimated across different subsets of samples if only the individual sparsity is imposed on the model. In this regard, applying a group sparsity on predictor variables is important, especially when the multivariate responses are presented.

3. The Proposed Bayesian Sparse MRMR

For the proposed BS-MRMR model, we make the following assumptions. First, assume the predictor variables are categorized into multiple groups and that a predictor is significant requires its variable group is significant. Second, assume that the distributions of response variables are from the exponential family. Third, assume that the hidden associations among mixed response variables can be represented in the precision matrix of latent Gaussian distributed variables. Now we present the details of the proposed BS-MRMR model.

Suppose that the predictor variables are $\boldsymbol{x} = (X_1, \ldots, X_p)^T$ and the multivariate mixed responses are $\boldsymbol{Y} = (\boldsymbol{U}^T, \boldsymbol{Z}^T, \boldsymbol{W}^T)^T$. Here $\boldsymbol{U} = (U^{(1)}, U^{(2)}, \ldots, U^{(l)})^T$ are the l-dimensional continuous responses, $\boldsymbol{Z} = (Z^{(1)}, Z^{(2)}, \ldots, Z^{(m)})^T$ are the m-dimensional counting responses, and $\boldsymbol{W} = (W^{(1)}, W^{(2)}, \ldots, W^{(k)})^T$ are the k-dimensional binary responses. To model the relationship between the predictor vector $\boldsymbol{x}$ and the response vector $\boldsymbol{Y}$, we consider a generalized linear model (GLM) for each individual response under appropriate link functions, while their link functions of the mean parameters form a multivariate linear model with respect to $\boldsymbol{x}$. Specifically,

$$
\begin{aligned}
U^{(j)}|\mu^{(j)}, \sigma^{(j)} &\sim N\left(\mu^{(j)}, \sigma^{(j)^2}\right), j = 1, \ldots, l,\\
Z^{(j)}|\lambda^{(j)} &\sim \text{Poisson}\left(\lambda^{(j)}\right), j = 1, \ldots, m,\\
W^{(j)}|\gamma^{(j)} &\sim \text{Bernoulli}\left(\gamma^{(j)}\right), j = 1, \ldots, k, \qquad (1)
\end{aligned}
$$

and

$$
\boldsymbol{\xi} = (\mu^{(1)}, \ldots, \mu^{(l)}, \log\lambda^{(1)}, \ldots, \log\lambda^{(m)}, \log\frac{\gamma^{(1)}}{1-\gamma^{(1)}}, \ldots, \log\frac{\gamma^{(k)}}{1-\gamma^{(k)}})^T \sim \mathcal{N}_q(\boldsymbol{B}^T\boldsymbol{x}, \boldsymbol{\Omega}^{-1}),
$$

where $\boldsymbol{B} = (\beta_{ij})_{p\times q}$ is a $p \times q$ coefficient matrix with $q = l + m + k$. Here $\mathcal{N}_q$ represents a q-dimensional multivariate normal distribution, and $\boldsymbol{\Omega}$ is the precision matrix of the error term $\boldsymbol{\varepsilon}$ by defining $\boldsymbol{\xi} = \boldsymbol{B}^T\boldsymbol{x} + \boldsymbol{\varepsilon}$. It is seen that the $\boldsymbol{\xi}$ is a latent vector connecting the multivariate mixed responses and predictor variables. The implication of $\boldsymbol{\Omega}$ is to characterize the conditional dependency relationship among the multivariate mixed responses $\boldsymbol{U}$, $\boldsymbol{Z}$, and $\boldsymbol{W}$. Denote the observed data as $(\boldsymbol{x}_i, \boldsymbol{y}_i), i = 1, \ldots, n$. Without loss of generality, we write $\boldsymbol{\xi}_i = \boldsymbol{B}^T\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i$ with $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_q(\boldsymbol{B}^T\boldsymbol{x}_i, \boldsymbol{\Omega}^{-1})$. Hence, the likelihood function can be expressed in a proper manner. To conduct the parameter estimation from a Bayesian perspective, we need to specify the priors for parameter matrices $\boldsymbol{B}$ and $\boldsymbol{\Omega}$, respectively.

3.1. Priors for Group and Individual Sparsity

Since the latent vector $\boldsymbol{\xi}$ follows the normal distribution, one can consider a conjugate prior of normal distribution for parameters $\text{vec}(\boldsymbol{B})$, where $\text{vec}(\cdot)$ is the vectorization operator. When fitting data with multivariate mixed responses, one would expect that only certain subgroups of predictor variables are related to the multivariate responses. Therefore, we would like to impose an appropriate prior for matrix $\boldsymbol{B}$ to enable variable selection in the sense that only a few groups of predictor variables are selected and a few coefficients are nonzeros within each selected group. Here we assume that the grouping of predictor variables is known. In particular, we propose to adopt a spike-and-slab prior (Liquet et al. 2017; Ning, Jeong, and Ghosal 2020) on the parameter matrix $\boldsymbol{B}$ for sparse estimation of parameters at both the group and individual levels.

Denote the data matrix by $\mathbb{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^T$. To facilitate the presentation, let the predictor vector $\boldsymbol{x} = (\boldsymbol{X}_1^T, \boldsymbol{X}_2^T, \ldots \boldsymbol{X}_G^T)^T$ to be composed of G groups with $\boldsymbol{X}_g$ containing p_g predictor variables for $g = 1, 2, \ldots, G$. Correspondingly, write $\mathbb{X} = (\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_G)^T$ and the coefficient matrix $\boldsymbol{B}$ is partitioned as $\boldsymbol{B} = (\boldsymbol{B}_1^T, \boldsymbol{B}_2^T, \ldots, \boldsymbol{B}_G^T)^T$, where $\boldsymbol{B}_g$ is a $p_g \times q$ matrix for the gth group of predictor variables. In order to enable variable selection for both group and individual levels, we re-parameterize the coefficients matrix in each group as $\boldsymbol{B}_g = \boldsymbol{V}_g\tilde{\boldsymbol{B}}_g$, where $\boldsymbol{V}_g = \text{diag}\{\tau_{g,1}, \ldots, \tau_{g,p_g}\}$ with $\tau_{g,j} \geq 0$ for $j = 1, \ldots, p_g$. The role of $\tau_{g,j}$ in diagonal matrix $\boldsymbol{V}_g$ is to control the sparsity of individual predictor variable within a group. That is, $\tau_{g,j} = 0$ corresponds to the jth predictor variable in the gth group being excluded from the regression model. Based on the above consideration, we employ the multivariate spike-and-slab prior for $\tilde{\boldsymbol{B}}_g$ as

$$
\begin{aligned}
\text{vec}(\tilde{\boldsymbol{B}}_g^T|\boldsymbol{\Omega}, \pi_1) &\sim (1-\pi_1)\mathcal{N}_{p_g q}(\boldsymbol{0}, \boldsymbol{I}_{p_g} \otimes \boldsymbol{\Omega}^{-1})\\
&\quad + \pi_1\delta_0(\text{vec}(\tilde{\boldsymbol{B}}_g^T)), \quad g = 1, \ldots, G \qquad (2)\\
\tau_{g,j}|\pi_2, \sigma_\tau^2 &\sim (1-\pi_2)\mathcal{N}^+(0, \sigma_\tau^2)\\
&\quad + \pi_2\delta_0(\tau_{g,j}), \quad g = 1, \ldots, G; j = 1, \ldots, p_g \qquad (3)\\
\pi_1 &\sim \text{Beta}(a_1, a_2), \ \pi_2 \sim \text{Beta}(a_3, a_4),\\
\sigma_\tau^2 &\sim \text{IG}(1, d),
\end{aligned}
$$

where $\boldsymbol{I}_a$ denotes the $a \times a$ identity matrix, the notation $\otimes$ stands for the Kronecker product, the symbol $\delta_0(\cdot)$ is the Dirac measure that denotes the point mass at 0, the symbol $\mathcal{N}^+(0, \sigma_\tau^2)$ represents a normal distribution $\mathcal{N}(0, \sigma_\tau^2)$ truncated below at 0, and IG(a, b) is the inverse Gamma distribution with its density function $f(x) = b^a x^{-(a+1)} \exp(-b/x)/\Gamma(a)$. The prior of $\tilde{\boldsymbol{B}}$ in (2) enables the variable selection at the group level, with our prior belief of the entire group $\boldsymbol{B}_g$ excluding from the model by the probability parameter π_1. Similarly, the prior of $\tau_{g,j}$ in (3) performs the variable selection at the individual level, with our prior belief of the jth row of $\boldsymbol{B}_g$ excluding from the model by the probability parameter π_2. Here we consider a Beta prior for π_1 to accommodate the potential domain-knowledge on the sparsity of the model. When there is no pre-knowledge on which group of predictor variables are related with responses, one could consider a simple uniform prior Unif(0,1) by setting $a_1 = a_2 = 1$. In this work, we adopt the suggestion in Scheipl, Fahrmeir, and Kneib (2012) to use an informative Beta prior Beta(20, 40), which is suitable for the high-dimensional data. Similarly, we adopt Beta(20, 40) as the prior of π_2 on the sparsity of individual predictor variable within each group. In addition, the parameter σ_τ^2 in the prior distribution of $\tau_{g,j}$ in (3) controls the shrinkage for the jth predictor variable in the gth group. A large value of σ_τ^2 may diffuse the coefficient for the corresponding predictor variable, and a small value may produce a biased estimated coefficient toward to zero. We thus use a conjugate inverse gamma prior IG$(1, d)$ for σ_τ^2 to determine its value from data, and adopt the "adaptive" idea for parameter d by estimating it with the Monte Carlo EM algorithm, which is proposed by Liquet et al. (2017). Specifically, in the kth EM iteration of estimating d, we update $d^{(k)} = \mathrm{E}^{-1}_{d^{(k-1)}}(1/\sigma_\tau^2|\text{rest})$, where the posterior expectation of σ_τ^2 is replaced by the Monte Carlo sample average of σ_τ^2 generated in the Gibbs samples based on $d^{(k-1)}$.

Next, we consider the prior of $\boldsymbol{\Omega}$ for inferring the conditional dependency relationship among responses. The conventional Bayesian methods for imposing sparsity on $\boldsymbol{\Omega}$ are implemented by the priors over the space of positive definite matrices constrained by fixed zeros. However, such priors often result in the daunting computational burdens for the large dimension of response variables. To address this challenge, we adopt the prior of $\boldsymbol{\Omega} = (\omega_{ij})_{q\times q}$ in (4), which is a spike-and-slab prior similar as Wang et al. (2015) for efficient computation.

$$\begin{aligned}\boldsymbol{\Omega}|\pi_3, \sigma_0^2, \sigma_1^2, \lambda \sim & \prod_{i<j}\big\{(1-\pi_3)\mathcal{N}(\omega_{ij}; 0, \sigma_0^2) \\ & +\pi_3\mathcal{N}(\omega_{ij}; 0, \sigma_1^2)\big\}\prod_i \mathrm{e}(\omega_{ii}; \frac{\lambda}{2})\boldsymbol{I}(\boldsymbol{\Omega} \in S^+) \\ \pi_3 \sim & \ \mathrm{Beta}(a_5, a_6), \end{aligned} \tag{4}$$

where $\boldsymbol{I}(\cdot)$ is the indicator function, S^+ stands for the cone of symmetric positive definite matrices, e$(\cdot)$ denotes the exponential distribution, and $\mathcal{N}(x; a, b)$ represents the density function of $\mathcal{N}(a, b)$ evaluated at point x. The term $(1-\pi_3)\mathcal{N}(\omega_{ij}; 0, \sigma_0^2)+\pi_3\mathcal{N}(\omega_{ij}; 0, \sigma_1^2)$ controls the sparsity on the off-diagonal elements ω_{ij}, and the term $\mathrm{e}(\omega_{ii}; \lambda/2)$ shrinkages the diagonal elements ω_{ii}. The prior of $\boldsymbol{\Omega}$ in (4) is computationally efficient since it can facilitate a fast block Gibbs sampler that updates the precision matrix $\boldsymbol{\Omega}$ one column at a time. While the conventional Bayesian methods update $\boldsymbol{\Omega}$ in a one-element-at-a-time manner. In practice, the value of σ_0^2 is set to be small, expressing our prior belief that the corresponding ω_{ij} is 0. On the other hand, the value of σ_1^2 is set to be large such that the estimated ω_{ij} would be very different from 0. Wang et al. (2015) demonstrated that when $\sigma_0 \geq 0.01$ and $\sigma_1/\sigma_0 \leq 1000$, the MCMC will converge quickly and mix quite well. Throughout this article in the numerical study, σ_0 and σ_1 are set to be 0.1 and 3, respectively. Here we consider an informative Beta prior for π_3 to encourage the sparsity in $\boldsymbol{\Omega}$ by setting $a_5 = q$ and $a_6 = q(q-1)/2$, given the prior belief that π_3 should be close to 0 to enhance the sparsity of $\boldsymbol{\Omega}$. We further consider to set hyperparameter $\lambda = q$ based on the observation from empirical studies that the structures of $\boldsymbol{\Omega}$ are insensitive to a range of λ. Note that Wang et al. (2015) also suggested to fix λ to 5 or 10 if the predictor variables are standardized.

3.2. Posterior and Inference

From Formula (1), we have the latent vector $\boldsymbol{\xi}_i = \boldsymbol{B}^T\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i$ with $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_q(\boldsymbol{B}^T\boldsymbol{x}_i, \boldsymbol{\Omega}^{-1})$. Let $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n)^T$ be an $n \times q$ matrix. Based on the priors above, the full-conditional distribution of unknown parameters conditional on the latent variable $\boldsymbol{\xi}$ and data is

$$\begin{aligned} & p(\tilde{\boldsymbol{B}}, \boldsymbol{\Omega}, \boldsymbol{\tau}, \pi_1, \pi_2, \pi_3, \sigma_\tau^2|\boldsymbol{\xi}) \\ \propto & p(\sigma_\tau^2)p(\pi_1)p(\pi_2)p(\pi_3)p(\tilde{\boldsymbol{B}}|\boldsymbol{\Omega}, \pi_1)p(\boldsymbol{\tau}|\pi_2, \sigma_\tau^2) \\ & p(\boldsymbol{\Omega}|\pi_3)\prod_{i=1}^n p(\boldsymbol{\xi}_i|\boldsymbol{X}, \tilde{\boldsymbol{B}}, \boldsymbol{\Omega}, \boldsymbol{\tau}). \end{aligned}$$

See the full expression in supplementary materials A.1. As a result, the full-conditional distribution of $\tilde{\boldsymbol{B}}_g$ is

$$\begin{aligned} \mathrm{vec}(\tilde{\boldsymbol{B}}_g^T|\text{rest}) \sim & (1-\pi_{B_g})\mathcal{N}_{p_g q}(\mathrm{vec}(\boldsymbol{M}_g^T), \boldsymbol{\Psi}_{p_g} \otimes \boldsymbol{\Sigma}) \\ & + \pi_{B_g}\delta_0(\mathrm{vec}(\tilde{\boldsymbol{B}}_g^T)) \end{aligned} \tag{5}$$

for $g = 1, \ldots, G$, where $\boldsymbol{\Psi}_{p_g} = (\boldsymbol{I}_{p_g} + \boldsymbol{V}_g\mathbb{X}_g^T\mathbb{X}_g\boldsymbol{V}_g)^{-1}$, $\boldsymbol{M}_g = \boldsymbol{\Psi}_{p_g}\boldsymbol{V}_g\mathbb{X}_g^T(\boldsymbol{\Xi} - \sum_{k\neq g}^G \mathbb{X}_k\boldsymbol{V}_k\tilde{\boldsymbol{B}}_k)$, and

$$\pi_{B_g} = \frac{\pi_1}{\pi_1 + (1-\pi_1)|\boldsymbol{\Psi}_{p_g}|^{\frac{q}{2}}\exp\left\{\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}\boldsymbol{M}_g^T\boldsymbol{\Psi}_{p_g}^{-1}\boldsymbol{M}_g)\right\}}.$$

Detailed derivation of π_{B_g} is provided in supplementary materials A.2. Denote by $\tilde{\boldsymbol{B}}_{gj}$ the jth row of $\tilde{\boldsymbol{B}}_g$, and $\mathbb{X}_{gj}$ the jth column of $\mathbb{X}_g$. Let $\boldsymbol{B}_{-gj}$ represent the matrix $\boldsymbol{B}$ without the jth row of group g, and $\mathbb{X}_{-gj}$ be the corresponding $\mathbb{X}$ without the jth column of group g. The full-conditional distribution of $\tau_{g,j}$ is

$$\tau_{g,j}|\text{rest} \sim (1-\pi_{\tau_{gj}})\mathcal{N}^+(\mu_{gj}, \sigma_{gj}^2) + \pi_{\tau_{gj}}\delta_0(\tau_{g,j}), \tag{6}$$

where $\sigma_{gj}^2 = [\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{B}}_{gj}^T\mathbb{X}_{gj}^T\mathbb{X}_{gj}\tilde{\boldsymbol{B}}_{gj}) + 1/\sigma_\tau^2]^{-1}$, $\mu_{gj} = \sigma_{gj}^2\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Xi}^T - \boldsymbol{B}_{-gj}^T\mathbb{X}_{-gj}^T)\mathbb{X}_{gj}\tilde{\boldsymbol{B}}_{gj}]$, and similarly we have

$$\begin{aligned}\pi_{\tau_{gj}} &= \frac{p(\tau_{g,j}=0|\text{rest})}{p(\tau_{g,j}=0|\text{rest}) + p(\tau_{g,j}\neq 0|\text{rest})} \\ &= \frac{\pi_2}{\pi_2 + 2(1-\pi_2)(\sigma_\tau^2)^{-\frac{1}{2}}(\sigma_{gj}^2)^{\frac{1}{2}}\exp\left\{\frac{1}{2}\frac{\mu_{gj}^2}{\sigma_{gj}^2}\right\}\Phi\left(\frac{\mu_{gj}}{\sigma_{gj}}\right)}, \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal variable.

The full-conditional distributions for π_1, π_2, π_3, and σ_τ^2 are

$$\pi_1|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{g=1}^{G} \boldsymbol{I}(\tilde{\boldsymbol{B}}_g = \boldsymbol{0}), a_2 + \sum_{g=1}^{G} \boldsymbol{I}(\tilde{\boldsymbol{B}}_g \neq \boldsymbol{0})\right), \tag{7}$$

$$\pi_2|\text{rest} \sim \text{Beta}\left(a_3 + \sum_{g=1}^{G}\sum_{j=1}^{p_g} \boldsymbol{I}(\tau_{g,j} = 0), a_4 + \sum_{g=1}^{G}\sum_{j=1}^{p_g} \boldsymbol{I}(\tau_{g,j} \neq 0)\right), \tag{8}$$

$$\pi_3|\text{rest} \sim \text{Beta}\left(a_5 + \sum_{i<j} \boldsymbol{I}(\omega_{ij} \neq 0), a_6 + \sum_{i<j} \boldsymbol{I}(\omega_{ij} = 0)\right), \tag{9}$$

$$\sigma_\tau^2|\text{rest} \sim \text{IG}\left(1 + \frac{1}{2}\sum_{g=1}^{G}\sum_{j=1}^{p_g} \boldsymbol{I}(\tau_{g,j} \neq 0), d + \frac{1}{2}\sum_{g=1}^{G}\sum_{j=1}^{p_g} \tau_{g,j}^2\right). \tag{10}$$

Next we examine the posterior distribution of $\boldsymbol{\Omega}$. To facilitate the expression, we introduce latent variables z_{ij} and re-write the prior (4) as

$$p(\boldsymbol{\Omega})| \propto \prod_{i<j} \mathcal{N}(\omega_{ij}; 0, \sigma_{z_{ij}}^2) \prod_{i}^{p} \text{e}(\omega_{ii}; \frac{\lambda}{2}),$$
$$p(z_{ij}) = \pi_3^{z_{ij}}(1-\pi_3)^{1-z_{ij}},$$

where $z_{ij} = 0$ or 1 according to whether $\omega_{ij} = 0$ or not. Hence, the variable z_{ij} follows Bernoulli distribution with parameter π_3. Now the full-conditional distribution of $\boldsymbol{\Omega}$ is

$$p(\boldsymbol{\Omega}|\text{rest}) \propto |\boldsymbol{\Omega}|^{\alpha/2} \exp\{-\frac{1}{2}\text{tr}(\boldsymbol{\Omega}\boldsymbol{\Theta})\} \prod_{i<j}\left\{\exp(-\frac{\omega_{ij}^2}{2\sigma_{z_{ij}}^2})\right\} \prod_{i}^{p} \exp\left(-\frac{\lambda}{2}\omega_{ii}\right), \tag{11}$$

$$p(z_{ij} = 1|\text{rest}) = \frac{\mathcal{N}\left(\omega_{ij}; 0, v_1^2\right)\pi_3}{\mathcal{N}\left(\omega_{ij}; 0, v_1^2\right)\pi_3 + \mathcal{N}\left(\omega_{ij}; 0, v_0^2\right)(1-\pi_3)},$$

where $\alpha = n + \sum_{g=1}^{G} p_g \boldsymbol{I}(\tilde{\boldsymbol{B}}_g \neq \boldsymbol{0})$ and $\boldsymbol{\Theta} = (\boldsymbol{\Xi} - \mathbb{X}\boldsymbol{B})^T(\boldsymbol{\Xi} - \mathbb{X}\boldsymbol{B}) + \tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}$. Sampling $\boldsymbol{\Omega}$ from its posterior (11) adopts the procedures described here:

Let $\boldsymbol{H} = (\sigma_{z_{ij}}^2)_{p\times p}$ be a symmetric matrix with zeros as its diagonal entries and $(\sigma_{z_{ij}}^2)_{i<j}$ as its upper diagonal entries. Partition $\boldsymbol{\Omega}$, $\boldsymbol{\Theta}$ and $\boldsymbol{H}$ as

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11}, & \boldsymbol{\varphi}_{12} \\ \boldsymbol{\varphi}_{12}', & \varphi_{22} \end{bmatrix}, \quad \boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_{11}, & \boldsymbol{\theta}_{12} \\ \boldsymbol{\theta}_{12}', & \theta_{22} \end{bmatrix}, \quad \text{and}$$
$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_{11}, & \boldsymbol{h}_{12} \\ \boldsymbol{h}_{12}', & 0 \end{bmatrix}.$$

Consider a variable transform: $(\boldsymbol{\varphi}_{12}, \varphi_{22}) \to (\boldsymbol{\eta} = \boldsymbol{\varphi}_{12}, \zeta = \varphi_{22} - \boldsymbol{\varphi}_{12}'\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\varphi}_{12})$. Then the conditional distributions of $\boldsymbol{\eta}$ and ζ are

$$\boldsymbol{\eta}|\text{rest} \sim \mathcal{N}(-\boldsymbol{\Sigma}_\eta\boldsymbol{\theta}_{12}, \boldsymbol{\Sigma}_\eta), \quad \text{and}$$
$$\zeta|\text{rest} \sim \text{Gamma}(\frac{\alpha}{2} + 1, \frac{\theta_{22} + \lambda}{2}),$$

where $\boldsymbol{\Sigma}_\eta = ((\theta_{22} + \lambda)\boldsymbol{\Omega}_{11}^{-1} + \text{diag}(\boldsymbol{h}_{12})^{-1})^{-1}$.

To construct matrix $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n)^T$ in the posteriors, we sample $\boldsymbol{\xi}_i$ according to

$$f(\boldsymbol{\xi}_i \mid \boldsymbol{y}_i, \boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Omega}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}[\boldsymbol{\xi}_i - \boldsymbol{B}^T\boldsymbol{x}_i]^T\boldsymbol{\Omega}[\boldsymbol{\xi}_i - \boldsymbol{B}^T\boldsymbol{x}_i]\right)$$
$$\prod_{j=1}^{l} \frac{1}{\sqrt{2\pi}\sigma^{(j)}} \exp\{-\frac{(u_i^{(j)} - \mu^{(j)})^2}{2\sigma^{(j)^2}}\}$$
$$\prod_{j=1}^{m} \frac{(\lambda^{(j)})^{z_i^{(j)}} \exp(-\lambda^{(j)})}{z_i^{(j)}!} \cdot \prod_{j=1}^{k} (\gamma^{(j)})^{w_i^{(j)}}(1-\gamma^{(j)})^{1-w_i^{(j)}}, \tag{12}$$

where a noninformative prior IG(1/2, 1/2) is assumed for the parameter $\sigma^{(j)^2}$, such that it can be sampled from the conditional distribution

$$\sigma^{(j)^2}|\text{rest} \sim \text{IG}\left(\frac{1}{2} + n, \frac{1}{2} + \frac{1}{2}\sum_{i=1}^{n}(u_i^{(j)} - \boldsymbol{\xi}_i^{(j)})\right), \tag{13}$$

where $\boldsymbol{\xi}_i^{(j)}$ is the jth element of $\boldsymbol{\xi}_i$. See a derivation of Equation (12) in supplementary materials A.3. Therefore, the Gibbs sampling for the proposed BS-MRMR model is summarized in Algorithm 1.

Algorithm 1 Gibbs sampling for BS-MRMR model

repeat
- Sample $\boldsymbol{\xi}_i$ by $f(\boldsymbol{\xi}_i \mid \boldsymbol{y}_i, \boldsymbol{B}, \boldsymbol{\Sigma})$ from (12) and (13).
- Sample $\tilde{\boldsymbol{B}}_g$ from (5).
- Sample $\tau_{g,j}$ from (6), and compute $\boldsymbol{B}_g$.
- Sample π_1, π_2, π_3 and σ_τ^2 from (7) to (10).
- Sample $\boldsymbol{\Omega}$ from (11).

until Convergence

4. Numerical Study

In this section, we evaluate the performance of the proposed model, denoted as BS-MRMR, by comparison with separate models (a) BS-GLM, (b) FS-GLM, and (c) a hierarchical generalized transformation (HGT) model (Bradley 2022). The BS-GLM method separately fits each response using the Bayesian generalized linear model, with its variable selection conducted according to the 95% credible intervals. The implementation of the BS-GLM method is conducted by *bayesglm*(·) function in the R software. The FS-GLM method separately fits each response on all the predictor variables via generalized linear model using the Lasso regularization. Precisely, the continuous, counting and binary responses are fitted through linear, Poisson and logistic regressions with Lasso penalties, respectively. This is implemented by *glmnet*(·) function in the R software. The third benchmark HGT model was recently developed to transform mixed responses into continuous responses (Bradley 2022). The HGT is a two-step model, where the first step samples latent continuous variables for the mixed responses, and the second step estimates a Bayesian multivariate regression model for each sample. The original article suggested to use a Bayesian mixed effects model as the second step estimator. However, such a model does not consider group and individual sparsity. For fair comparison, a sparse multivariate regression model called

MBSGSSS (Liquet et al. 2017) was selected to be the second step model here. In the rest of this manuscript, we denote the HGT-MBSGSSS model as HGT for short.

Regarding the dependency among multiple responses, we consider the following matrix structures of $\boldsymbol{\Omega} = (\omega_{ij})_{q\times q}$.

- Scenario 1 $\boldsymbol{\Omega}_1$: $\omega_{ij} = 1_{\{i=j\}} + 0.5_{|i-j|=1} + 0.3_{|i-j|=2} + 0.1_{|i-j|=3}$.
- Scenario 2 $\boldsymbol{\Omega}_2$ is generated by randomly permuting rows and corresponding columns of $\boldsymbol{\Omega}_1$.
- Scenario 3 $\boldsymbol{\Omega}_3 : \omega_{ij} = 0.5^{|i-j|}$.
- Scenario 4 $\boldsymbol{\Omega}_4$: randomly and evenly divide the indices $1, 2, \ldots, q$ into M groups. Let $\omega_{jj} = 1$. Set $\omega_{jk} = 0.4$ for $j \neq k$ if j and k belong to the same group and 0 otherwise.
- Scenario 5 $\boldsymbol{\Omega}_5 = \begin{pmatrix} \text{CS}(0.5) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{pmatrix}$, where CS(0.5) represents a 4×4 compound structure matrix with diagonal elements 1 and others 0.5. $\mathbf{0}$ indicates a matrix with all elements 0.

Scenario 1 is a banded matrix representing each response is only correlated with its several nearest responses. Scenario 2 disrupts such sparse structure but maintaining the same sparse extent. Scenario 3 is an autoregressive model with its elements decaying as one moves away from the diagonal. Scenario 4 is a random sparse matrix. Scenario 5 represents that the first 4 responses are correlated but independent from others. We generate $n = 100$ training data to estimate the model and 100 testing data to examine the model performance, both of which are from the multivariate normal distribution $\mathcal{N}_p(0, \sigma_X \boldsymbol{I})$ with (a) $p = 20, l = m = k = 2$ and (b) $p = 80, l = m = k = 5$. Here we choose $n = 100$ as the training sample size to stress the proposed model with a large number of parameters in comparison with the sample size. For example, $p = 20, l = m = k = 2$ results in $(p+1)(l+m+k) = 126$ linear model coefficients to be estimated including the intercepts; and $p = 80, l = m = k = 5$ results in $(p+1)(l+m+k) = 1215$ linear model coefficient parameters. We also we consider a simulation setting with $n = 50, p = 80$ to further evaluate the proposed model. The results are summarized in the supplementary materials (Tables 5–7).

When $p = 20$, we take $M = 3$ for Scenario 4, and the coefficient matrix is divided into four groups with each group having five variables as $\boldsymbol{B}_1^T = (** 00 * \mathbf{0}_5 ** 00 * \mathbf{0}_5)$ and $\boldsymbol{B}_2^T = (\mathbf{0}_5 ** 00 * ** 00 * \mathbf{0}_5)$, where $*$ represents that the corresponding columns are not zeros generating from uniform distribution $\text{Unif}(l_B, u_B)$, 0 represents that the corresponding columns are zeros, and $\mathbf{0}_a$ represents that the corresponding a columns are zeros. When $p = 80$, the value of M is set to be 5 in Scenario 4. We consider 6 groups of predictor variables and divide coefficient matrix $\boldsymbol{B}_3^T = ((\mathbf{0}_5, *_5)\ \mathbf{0}_{20}\ (*_5, \mathbf{0}_5)\ \mathbf{0}_{10}\ (*_5, \mathbf{0}_{15})\ \mathbf{0}_{10})$ as well as $\boldsymbol{B}_4^T = (\mathbf{0}_{20}\ (\mathbf{0}_5, *_5)\ (*_5, \mathbf{0}_5)\ (*_5, \mathbf{0}_{15})\ \mathbf{0}_{10}\ \mathbf{0}_{10})$. The observations of the response variables are then generated based on Formula (1). The data generation parameters σ_X, l_B and u_B are tuned to make sure that the counting observations in the response matrix for each setting are within a reasonable range. We generate 10,000 MCMC samples with the first 2000 draws as burn-in period. The median values of the rest 8000 samples are taken as the parameter estimates for the proposed model.

To evaluate the accuracy of model estimation, we consider the loss measures

$$L(\hat{\boldsymbol{B}}) = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(\boldsymbol{B}_{ij} - \hat{\boldsymbol{B}}_{ij})^2}{pq}} \quad \text{and}$$

$$L(\hat{\boldsymbol{\Omega}}) = \sqrt{\frac{\sum_{i=1}^{q}\sum_{j=1}^{q}(\boldsymbol{\Omega}_{ij} - \hat{\boldsymbol{\Omega}}_{ij})^2}{q^2}},$$

where $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{\Omega}}$ denote the estimates of matrices $\boldsymbol{B}$ and $\boldsymbol{\Omega}$. For gauging the performance of variable selection in $\hat{\boldsymbol{B}}$ and sparsity imposed in $\hat{\boldsymbol{\Omega}}$, we also consider FSL = false positive (FP) + false negative (FN). Figures 1 and 2 summarize via boxplots the results of two selected scenarios of these loss measures for each method over 50 replicates. Note that parameter estimation errors of the HGT method are not reported in Figure 1 since the HGT method transforms responses into different scales, making it noncomparable with other methods. Please refer to Tables 1 and 2 in supplementary materials for the full comparison results.

It is clear to see that the proposed BS-MRMR model generally outperforms other compared methods for all the settings with respect to loss measures $L(\hat{\boldsymbol{B}})$, $L(\hat{\boldsymbol{\Omega}})$, $FSL(\hat{\boldsymbol{B}})$, and $FSL(\hat{\boldsymbol{\Omega}})$. The proposed model produces the lowest values of losses and corresponding standard errors, because it takes advantage of the association between responses and model them jointly. Similar conclusion can be readily drawn by comparing the HGT model with two separate GLM benchmark models. The main reason is that the BS-GLM and FS-GLM methods fit data separately, losing the potential information of the response variables' relationship. Specifically, when the number of predictor variables $p = 20$, the BS-GLM method sometimes performs better than the FS-GLM for the loss $L(\hat{\boldsymbol{B}})$, but other times it is worse, depending on the structures of $\boldsymbol{\Omega}$ and coefficient matrices $\boldsymbol{B}$. When p increases to 80, the FS-GLM method appears to be better regarding the loss $L(\hat{\boldsymbol{B}})$. Nevertheless, they are all inferior to the proposed model. For the loss $L(\hat{\boldsymbol{\Omega}})$, note that the separate modeling FS-GLM cannot provide the correlation among responses. And the metric $FSL(\hat{\boldsymbol{\Omega}})$ for HGT is not accessible since the function "MBSGSSS" provided by Liquet et al. (2017) does not return samples for $\hat{\boldsymbol{\Omega}}$. In addition, regarding the loss $FSL(\hat{\boldsymbol{B}})$, the proposed BS-MRMR model is substantially superior over the compared methods, implying that it is able to accurately identify the group sparsity and individual sparsity within each group. We also observe that the BS-GLM method performs better than the FS-GLM with respect to $FSL(\hat{\boldsymbol{B}})$. It is also interesting to observe that the HGT model can have a comparable performance with the proposed BS-MRMR model in terms of $FSL(\hat{\boldsymbol{B}})$. Note that, in the current implementation of the HGT method, we adopted the default settings of hyperparameters in the supplemented programming code in Bradley (2022). When comparing the loss $FSL(\hat{\boldsymbol{\Omega}})$, the proposed BS-MRMR model does not consistently outperform the BS-GLM, especially when the number of responses increases from 6 to 15. The reason is that the BS-GLM only provides estimates for diagonal elements in $\hat{\boldsymbol{\Omega}}$, hence, the loss $FSL(\hat{\boldsymbol{\Omega}})$ tends to be smaller as the underlying $\boldsymbol{\Omega}$ becomes sparser.

To further investigate the prediction performance of the proposed model, the root-mean-square error (RMSE) $\sqrt{\sum_{i=1}^{100}(y_i - \hat{y}_i)^2/100}$ is computed for the continuous and counting responses via the testing data, denoted by RMSE(N)

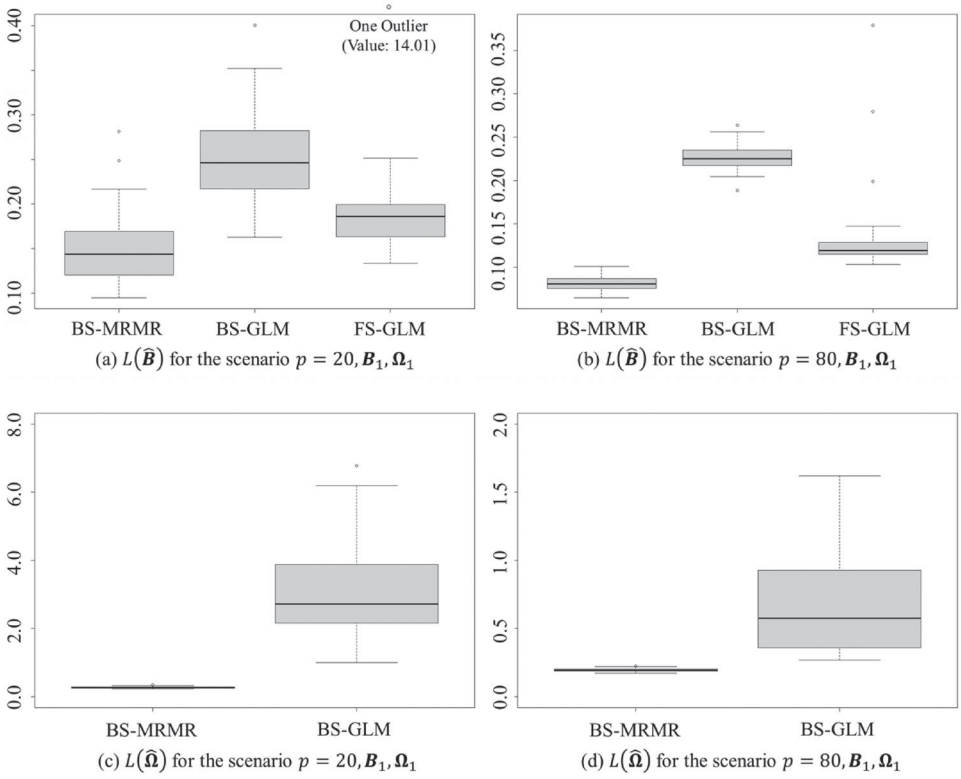

Figure 1. Boxplots of two selected scenarios for parameter estimation errors of $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{\Omega}}$ to compare BS-MRMR (proposed), FS-GLM, and BS-GLM. See full results in supplementary materials (Table 1).

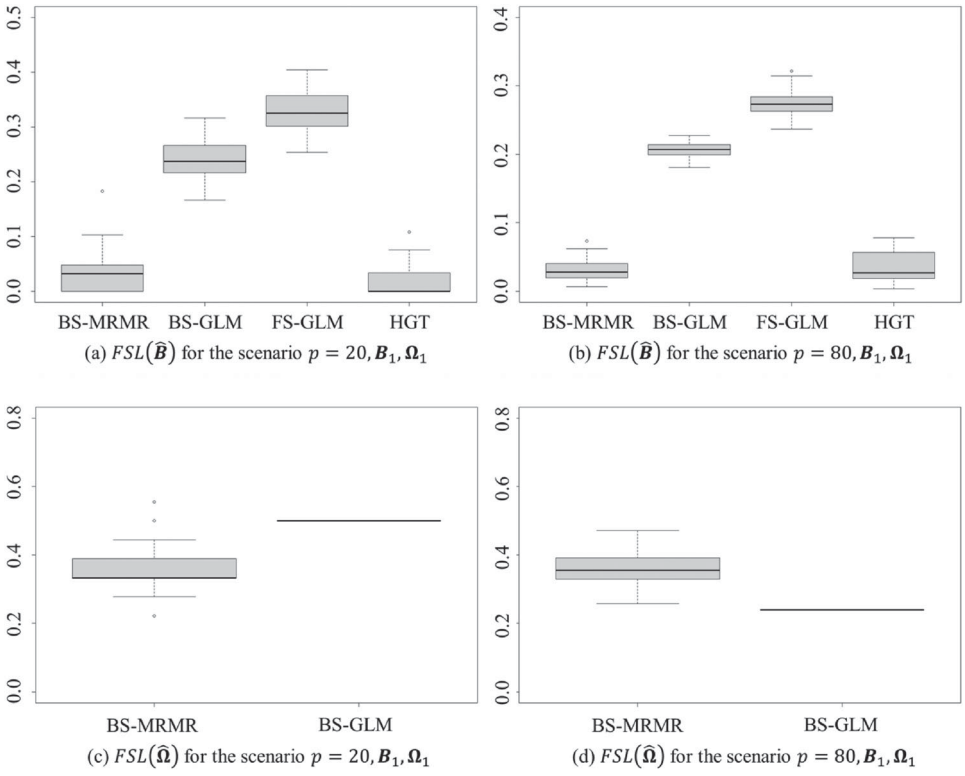

Figure 2. Boxplots of two selected scenarios for variable selection errors of $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{\Omega}}$ to compare BS-MRMR (proposed), FS-GLM, BS-GLM, and HGT. See full results in supplementary materials (Table 2).

and RMSE(P), where $\hat{y}_i$ stands for the corresponding fitted value. Naturally, we use the misclassification error rate (ME) $\sum_{i=1}^{100} \boldsymbol{I}(y_i \neq \hat{y}_i)/100$ to compare the model performance on the binary response. The cutoff point for the binary response estimates is 0.5. Figure 3 reports the prediction results of the selected scenarios for the methods in comparison. Please see the

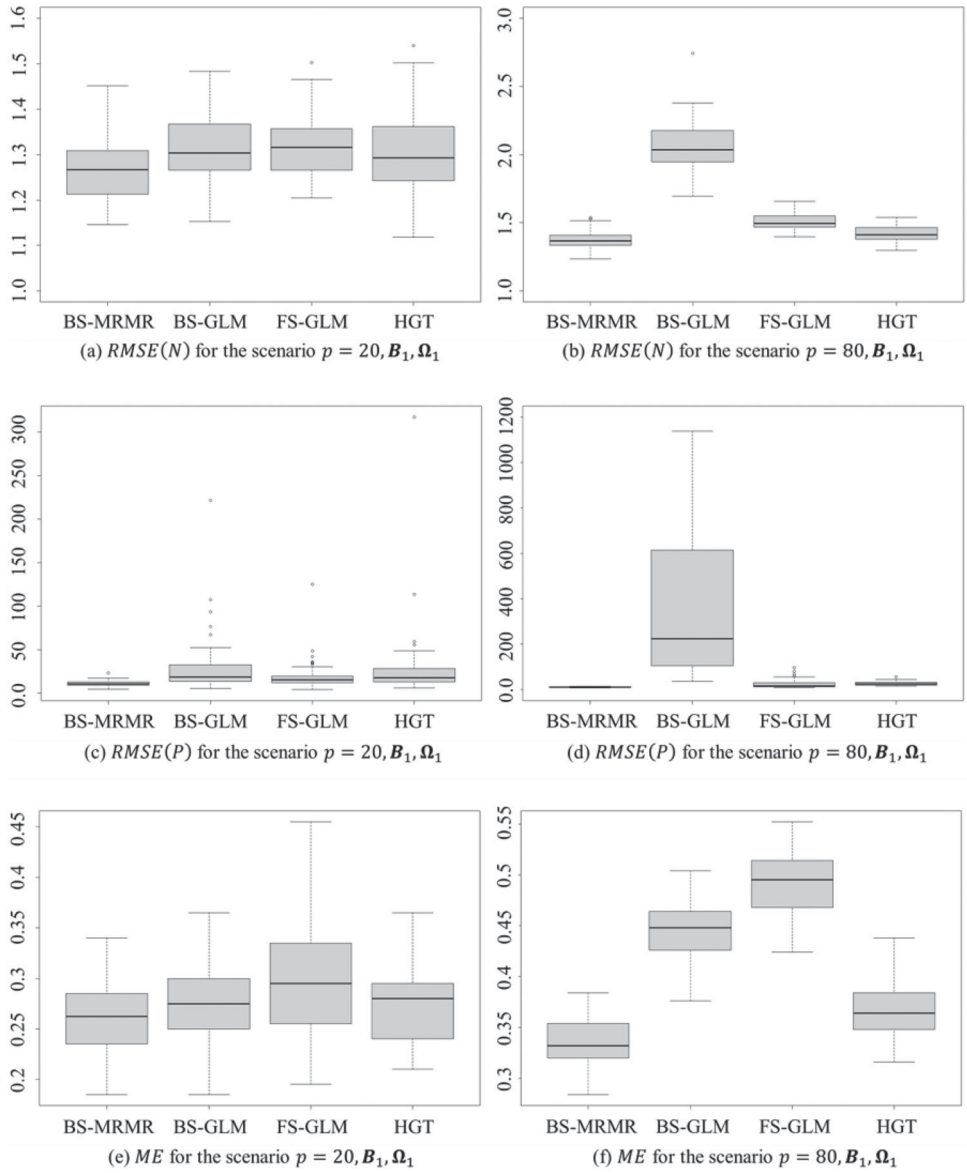

Figure 3. Boxplots of two selected scenarios for prediction errors to compare BS-MRMR (proposed), FS-GLM, BS-GLM, and HGT. See full results in supplementary materials (Tables 3 and 4).

full comparison results in the Tables 3 and 4 in supplementary materials. We observe that the proposed BS-MRMR model provides more accurate predictions on continuous, counting and binary responses for all the settings. Specifically, it is slightly better than the compared methods with respect to RMSE(N), and substantially better in terms of RMSE(P) and ME. When the number of predictor variables increases from 20 to 80, the proposed model performs consistently well with the lowest loss values and standard errors. Such results demonstrate that incorporating the association of responses in the proposed model will obviously improve its prediction performance. In addition, we also observe that the FS-GLM and BS-GLM methods occasionally have convergence issue when fitting the counting responses, yielding high values of RMSE(P) especially for larger $p = 80$. This is possibly due to some very large values of counting responses in the datasets, which remarkably increases the difficulty of modeling and hence causes convergence issue. In contrast, the relatively lower values of RMSE(P) produced by the proposed method demonstrate that the BS-MRMR model is more robust than the compared approaches for the underlying multivariate datasets.

Figure 4 shows the trace plots of randomly selected parameters in the precision matrix $\boldsymbol{\Omega}$ and coefficient matrix $\boldsymbol{B}$ from one replicate for Scenario 1 and $\boldsymbol{B}_3$ when $p = 80$. It is seen that the traces of the parameters fluctuate around the means with relatively stable variation, indicating that the MCMC chains converge. Figure 5 displays the corresponding autocorrelation functions of those parameters. The quick decrease of ACF in these plots implies the fast convergence of the Gibbs sampling iterations. The rest of the parameters present the similar patterns, and hence their plots are omitted. We further compare the computation time for each method for the simulation study ($p = 20, \boldsymbol{B}_1, \boldsymbol{\Omega}_1$). The average time for model estimation and prediction are summarized in Table 1. We note that the long computation time of HGT could be due to the current implementation of the two-step estimation procedure. Specifically, the implementation of HGT method first samples latent continuous

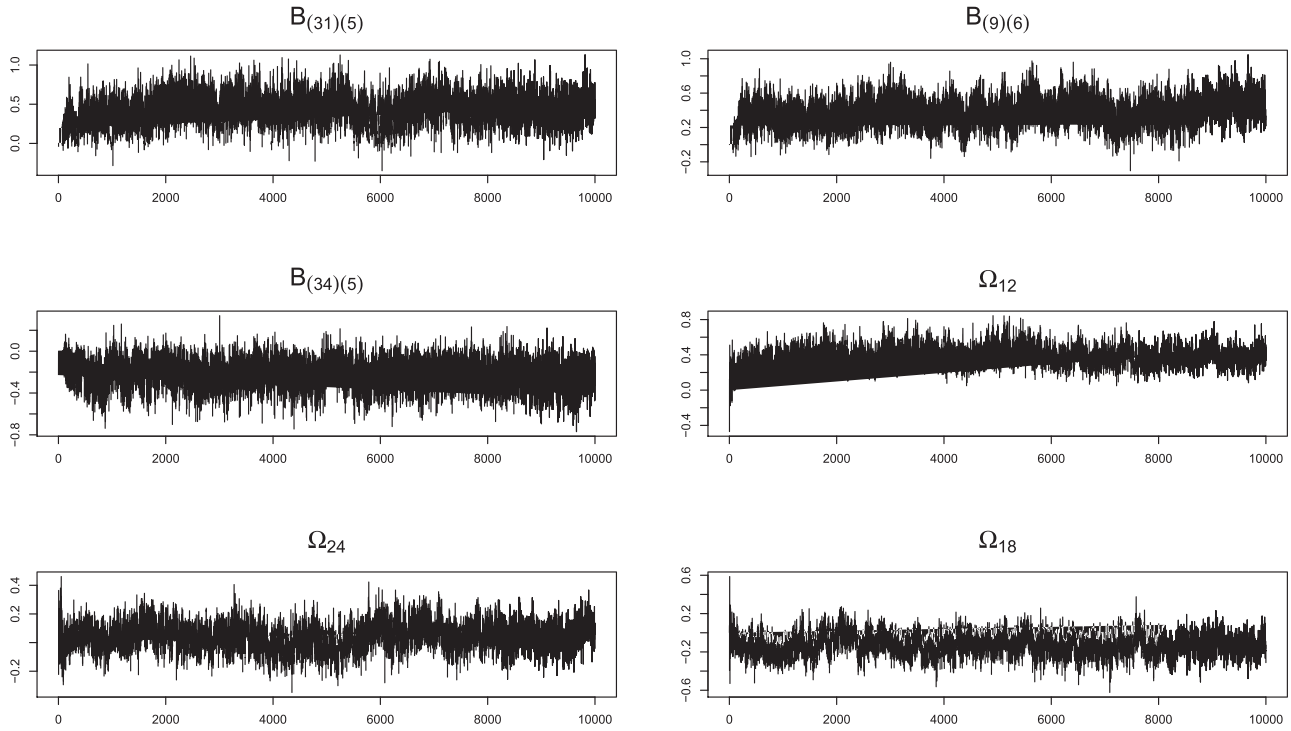

Figure 4. Trace plots for selected parameters from one replicate for Scenario 1 and $\boldsymbol{B}_3$ when $p = 80$.

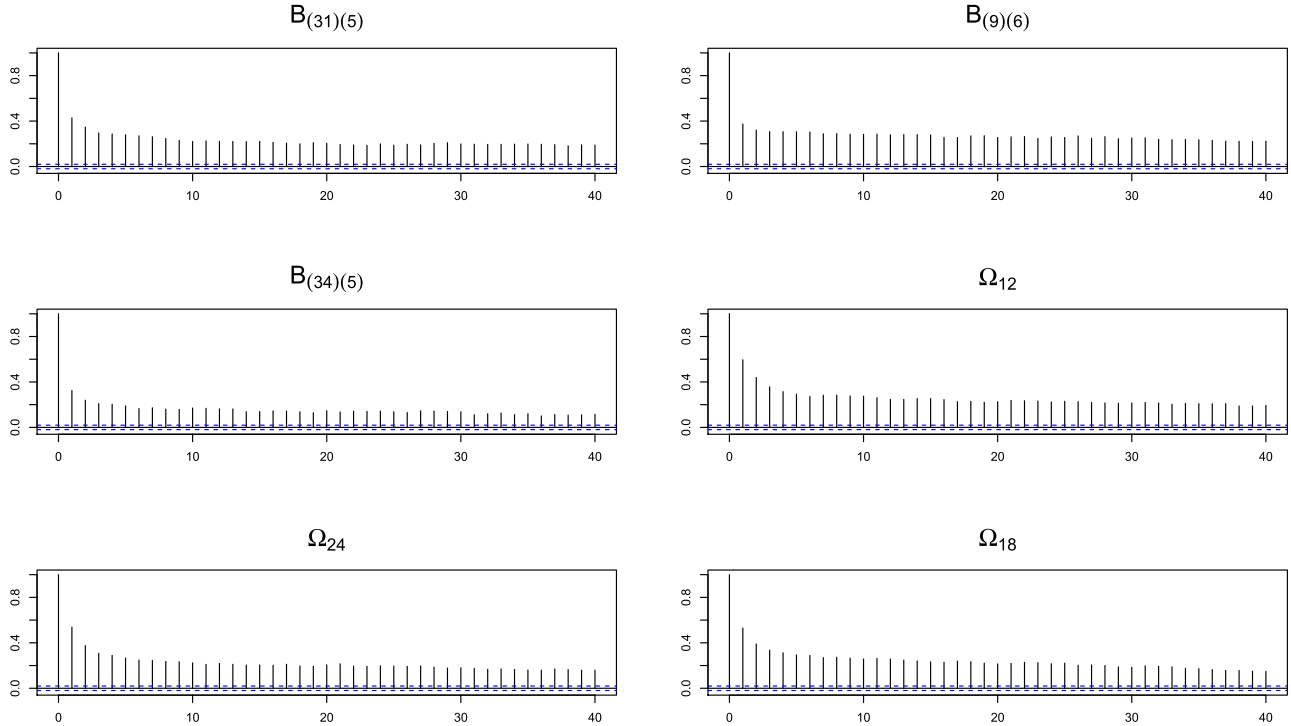

Figure 5. ACF plots for selected parameters from one replicate for Scenario 1 and $\boldsymbol{B}_3$ when $p = 80$.

variables for the mixed responses, then estimates a Bayesian multivariate regression model for each sample. Consequently, the estimation time of HGT is linearly related to the number of samples obtained from the first step. It is worth to remarking that there could be a computationally more efficient implementation by combining the Gibbs sampler of HGT with the second step sampler (Bradley 2022).

Table 1. The averages and standard errors (in parenthesis) of computation time for the numerical study with $p = 20, \boldsymbol{B}_1, \boldsymbol{\Omega}_1$.

	Estimation time (sec)	Prediction time (sec)
FS-GLM	334.9 (5.854)	0.158 (0.004)
BS-GLM	0.035 (0.000)	0.001 (0.000)
HGT	67652 (108.5)	4.414 (0.041)
BS-MRMR	0.059 (0.001)	0.002 (0.000)

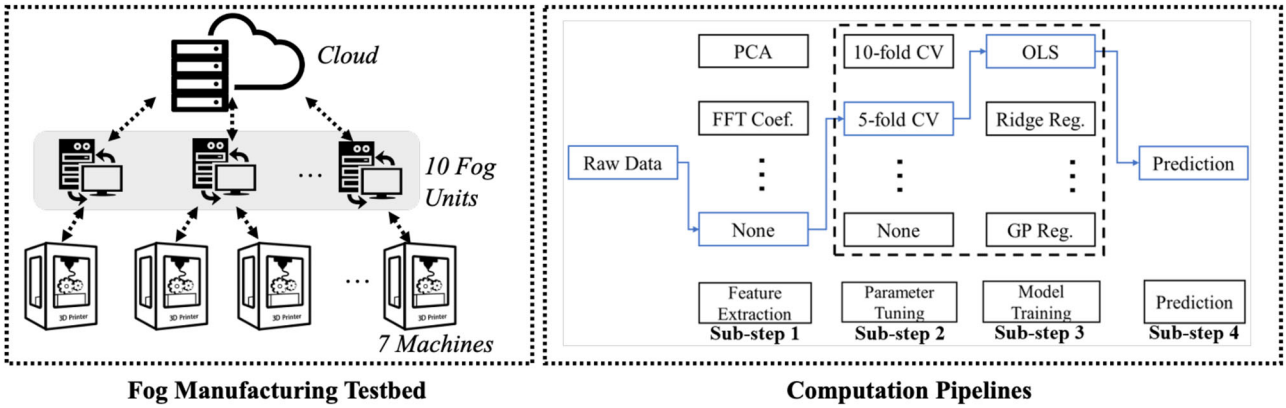

Figure 6. Fog manufacturing testbed and recommended computation pipelines. Redrawn from Zhang et al. (2019) with authors' permission.

Table 2. Design of experiments for the analysis of runtime performance metrics (summarized from Zhang et al. (2019) with authors' permission).

Design factors	Level 1	Level 2
Task selection strategy	Random selection	Recommendation (Chen and Jin 2020)
Number of pipelines	5	10
Data storage strategy	One copy on each Fog node	Three copies randomly stored on three Fog nodes
Offloading strategy	Random offloading	Time-balanced offloading

5. Case Study in Fog Manufacturing

This section investigates a real case study in a Fog manufacturing testbed for evaluating the proposed BS-MRMR model. The three-layer architecture of this testbed is presented in the left panel of Figure 6. In the top layer, a central computation unit (CPU: i7-6700k) serves as the orchestrator to master the offloading of computation tasks and collect the results from each Fog unit. In the middle layer, 10 Raspberry Pi 3 devices are deployed as Fog nodes with different computation capabilities and communication bandwidths. In the bottom layer, seven manufacturing machines are connected to each Fog nodes, such that the manufacturing process data from any machine can be collected by each of 10 Fog nodes. All the runtime performance signals (i.e., CPU and memory utilizations, temperature of the Fog nodes, etc.) during the execution are stored locally in each Fog nodes based on a Python program. For the computation tasks to be offloaded, the right panel of Figure 6 presents the computation pipelines with four sub-steps and multiple method options in each sub-step following the definition described in Chen and Jin (2020). Here, each option from one sub-step is treated as a computation task, and one sequence from sub-step 1 to sub-step 4 represents a computation pipeline. The predictive offloading methods aim at dynamically assigning these computation tasks into different Fog nodes considering the responsiveness and reliability as detailed in Chen et al. (2018). The offloading decisions are made based on the prediction of the runtime performance metrics extracted from the runtime performance signals.

To generate the runtime performance signals for analysis, an experiment of four factors with two levels was conducted by the full factorial design in Table 2. There are 32 runs in total executed with two replicates for each treatment. The workflow of this Fog manufacturing testbed in this experiment follows three steps: first, the computation pipelines to be offloaded in sub-steps are selected following two task selection strategies, namely, (a) random selection from all candidate pipelines, and (b) recommendation-based selection to choose the Top-ranked pipelines suggested by a recommender system (see Chen and Jin 2020 for details). Second, the orchestrator then provides the offloading decisions (randomly or following a time-balanced offloading strategy) to assign sub-steps of the selected computation pipelines to different Fog nodes for execution. Finally, Fog nodes check whether the dataset to support the assigned sub-steps exists in their local storage according to the data storage strategy. A Fog unit will download the necessary dataset from other Fog nodes or the orchestrator, then will execute the assigned sub-step with runtime performance metrics recorded.

In this case study, the observational data are $\{\boldsymbol{x}_{t_f}, \boldsymbol{Y}_{t_f}\}$, where $t_f = 1, \ldots, n_f$ is defined as the t_fth sub-step that is assigned to the fth Fog unit, $f = 1, \ldots, 10$. The $\boldsymbol{x}_{t_f^{(1)}} \in \mathbb{R}^{48}$ is the predictor vector that contains two groups of features (i.e., Group 1: 11 summary statistics of each of the three runtime performance signals; and Group 2: 17 dummy variables as the embedding of the assigned sub-step) from a previous time instance. The $\boldsymbol{Y}_{t_f^{(2)}} \in \mathbb{R}^5$ is the response vector which contains five runtime performance metrics in mixed types when executing the t_fth sub-step, that is, continuous metrics: averaged CPU utilization $\mathbb{Y}_1$, averaged temperature $\mathbb{Y}_2$; counting metric: number of sub-steps that can be executed within 5 sec $\mathbb{Y}_3$; and binary metrics: whether temperature exceeds a certain threshold $\mathbb{Y}_4$, and whether memory utilization exceeds a certain threshold $\mathbb{Y}_5$. In total, 3407 samples were obtained from the total 10 Fog nodes. Each Fog unit has around 340 samples ordered by timestamps when the samples are collected. The training data consists of the first 200 samples from each Fog unit, and the testing data consists of the remaining samples from each Fog unit.

We compare the proposed BS-MRMR model with the FS-GLM, BS-GLM, and HGT model. For the BS-MRMR, the values of hyperparameters of priors are the same as those in Section 4 to encourage both group and individual sparsity for predictor variables. In addition, 10,000 MCMC samples from the proposed model are drawn with the first 2000 as burn-in period. For the HGT method, we adopted the default settings

Table 3. The averages and standard errors (in parenthesis) of loss measures for the real case study in Fog manufacturing.

	RMSE(N)	RMSE(P)	ME
FS-GLM	1.345(0.054)	102.9(45.57)	0.058(0.024)
BS-GLM	6.074(1.344)	338.0(52.65)	0.240(0.055)
HGT	3.186(0.366)	20.79(5.723)	0.263(0.023)
BS-MRMR	0.521(0.044)	10.24(0.434)	0.039(0.012)

of hyperparameters in the supplemented programming code in Bradley (2022). For this real data of Fog manufacturing, there can be 12 hyperparameters in the HGT method. It would be very challenging to carefully adjust hyperparameters to avoid the bias issue introduced by transformation. One possibility may specify hyperparameters for the HGT methods using certain Bayesian optimization techniques when the number of hyperparameters is large.

The averages and standard errors of loss measures RMSE(N), RMSE(P), and ME over all 10 Fog nodes are summarized in Table 3. It can be readily observed that the proposed BS-MRMR model consistently outperforms the frequentist and Bayesian separate models, and HGT model for all types of responses. The BS-MRMR performs significantly the best in the prediction of counting runtime performance metric, which may be attributed to the shared information from other correlated metrics. We then further investigate the estimated precision matrix $\hat{\boldsymbol{\Omega}}$ and the corresponding correlation matrix. In Figure 7(a), the median values of $\hat{\boldsymbol{\Omega}}$ and the estimated 95% credible intervals are visualized in matrix bar plots with error bars. Note that the median values of $\hat{\boldsymbol{\Omega}}$ are standardized in the range of $[-1, 1]$ with diagonal elements to be all ones. Figure 7(b) plots the median values of correlation matrix with 95% credible intervals, which are converted from the estimated precision matrix. This correlation matrix well aligns with the generation of the runtime performance metrics. For example, the counting metric $\mathbb{Y}_3$ is correlated with the continuous metric $\mathbb{Y}_1$ (i.e., $\text{corr}(\xi_1, \xi_3) = 0.218$), since the number of sub-steps that can be executed in 5 sec highly depends on the CPU utilization. As another example, $\mathbb{Y}_2$ is highly correlated with $\mathbb{Y}_4$ (i.e., $\text{corr}(\xi_2, \xi_4) = 0.637$), since $\mathbb{Y}_4$ is generated by comparing $\mathbb{Y}_2$ with a certain threshold. Besides, Figure 7 presents the sparsity of the estimated precision matrix $\hat{\boldsymbol{\Omega}}$ with narrow credible intervals, which demonstrates the effectiveness of the slack-and-slab prior imposed in the precision matrix $\boldsymbol{\Omega}$.

Moreover, we also investigate statistical inferences for the uncertainty quantification of the predicted mixed metrics. Figure 8(a) reports the median of the predicted latent responses $\xi_1, \ldots, \xi_5$ by the BS-MRMR and the associated 95% credible intervals on the testing data. In Figure 8, 95% credible intervals are presented by the shaded region, the median values of the predicted latent variables are plotted in black solid lines, and the true responses are plotted in blue dotted lines. Figure 8(b) and (c) report the prediction and the associated 95% credible intervals on the testing data from the BS-GLM and HGT methods, respectively. It is noted that the true responses are mainly contained by the 95% credible intervals from the proposed BS-MRMR model, while the 95% credible intervals of the BS-GLM and HGT do not perform well to cover the true responses. The narrow credible intervals of the BS-MRMR model can

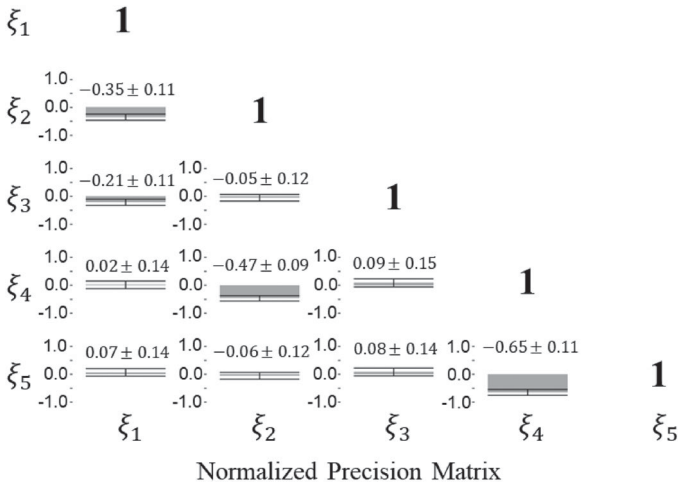

Figure 7. Estimated dependency (i.e., precision matrix) of the latent responses (i.e., $\xi_1, \ldots, \xi_5$) from the BS-MRMR method, where the shadow bars present the median values, and 95% credible intervals are presented in the form of error bars and labels on subplots.

Table 4. Settings for sensitivity study with five factors and two levels at each factor.

Factors	Level 1	Level 2
(a_1, a_2)	$(1, 1)$	$(2, 2)$
(a_3, a_4)	$(2p, p)$	(p, p)
(a_5, a_6)	$(q, q(q-1)/2)$	(q, q)
(α, λ)	$(q/2, q)$	(q, q)
(σ_0, σ_1)	$(0.1, 3)$	$(0.2, 2)$

be attributed to the joint modeling of mixed responses and the quantification of hidden associations among these mixed responses. Note that the narrow credible intervals indicate low uncertainty in predicting runtime performance metrics. In addition, it is seen that the BS-GLM with Poisson response provides unstable prediction, which leads to extremely large credible intervals (see $\exp(\boldsymbol{\xi}_3)$ in Figure 8(b)). Besides, FS-GLM cannot provide uncertainty quantification, hence, its intervals are not available.

In addition, we also conduct the sensitivity analysis on the choice of priors with respect to hyper-parameters in the proposed BS-MRMR model. Specifically, we set a full factorial design for five pairs of hyperparameters as five factors with two levels for each factor. The five factors and levels are listed in Table 4. The 10-fold cross-validations are used to check the prediction performance of the proposed method under $2^5 (= 32)$ setting of hyper-parameters. Detailed settings for 32 designs are summarized in Table 8 of the supplementary materials. From the results (Table 5), one can see there is not significant differences under different setting of prior hyper-parameters, indicating that the proposed BS-MRMR method is not sensitive to the choice of priors.

To support predictive offloading in Fog manufacturing, the offloading method should determine both the offloading strategies (e.g., randomly offloading, closest distance-based offloading, etc.) and the offloading decisions by considering not only the predicted runtime performance metrics, but also the uncertainty associated with the predictions. For example, it is confident for the predictive offloading strategy to optimize the

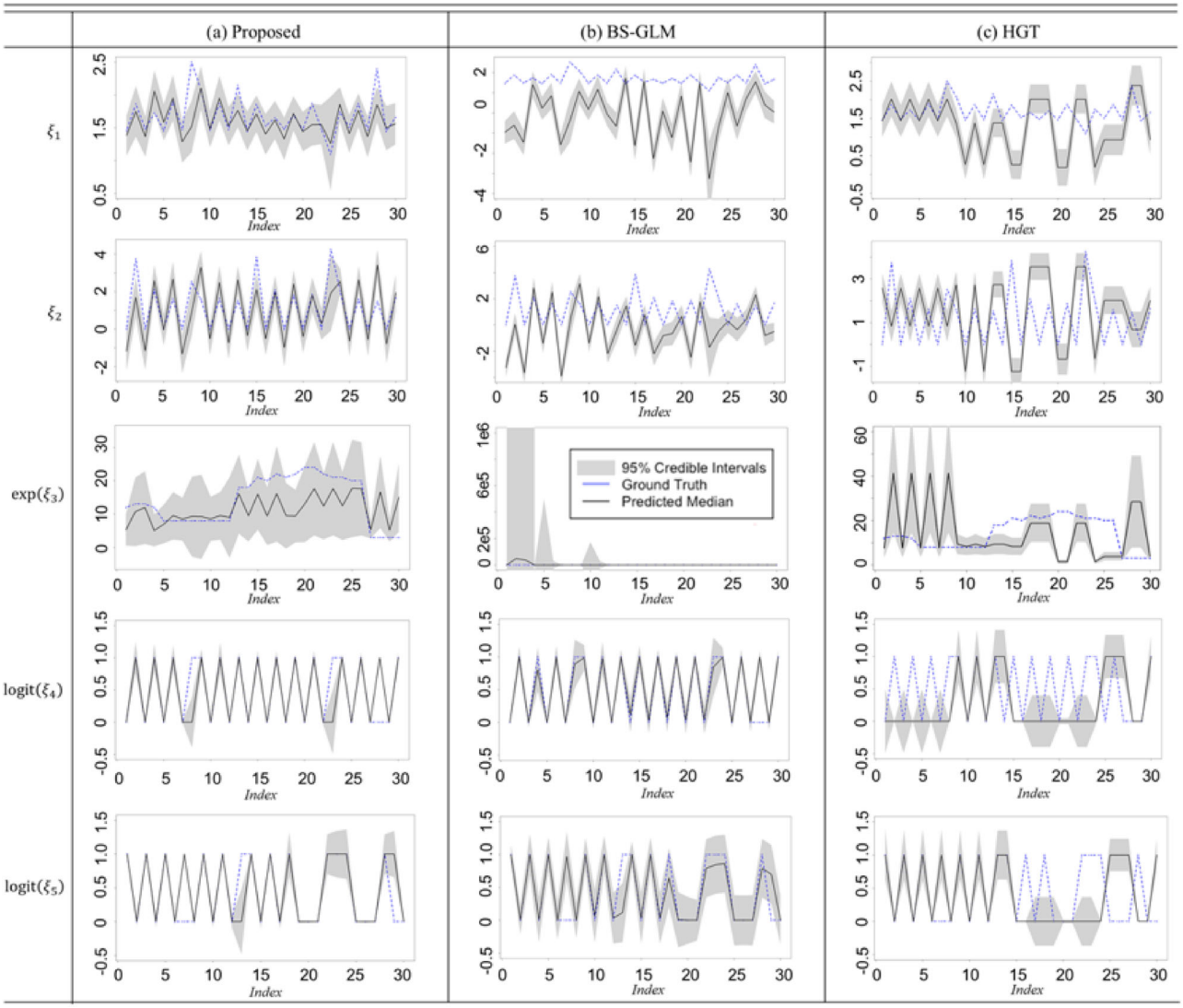

Figure 8. The median of the predicted latent responses $\xi_1, \ldots, \xi_5$ and the associated 95% credible intervals on the testing data: (a) the proposed BS-MRMR, (b) the BS-GLM, (c) the HGT.

Table 5. Sensitivity study results to explore different combinations of priors.

DOE	RMSE Normal	RMSE Poisson	ME	DOE	RMSE Normal	RMSE Poisson	ME
1	0.515(0.043)	10.281(0.417)	0.051(0.010)	17	0.517(0.044)	10.248(0.459)	0.053(0.011)
2	0.516(0.044)	10.141(0.418)	0.055(0.011)	18	0.510(0.043)	10.234(0.418)	0.054(0.011)
3	0.520(0.044)	10.325(0.487)	0.055(0.011)	19	0.521(0.044)	10.385(0.447)	0.051(0.010)
4	0.519(0.044)	10.227(0.406)	0.056(0.011)	20	0.519(0.043)	10.490(0.494)	0.052(0.011)
5	0.514(0.044)	10.341(0.523)	0.050(0.010)	21	0.525(0.044)	10.239(0.437)	0.053(0.011)
6	0.515(0.044)	10.291(0.445)	0.051(0.010)	22	0.517(0.044)	10.224(0.442)	0.055(0.011)
7	0.522(0.045)	10.303(0.444)	0.055(0.010)	23	0.518(0.044)	10.406(0.441)	0.053(0.011)
8	0.524(0.044)	10.367(0.483)	0.055(0.012)	24	0.522(0.044)	10.170(0.447)	0.052(0.010)
9	0.517(0.044)	10.289(0.422)	0.054(0.010)	25	0.529(0.044)	10.318(0.424)	0.055(0.010)
10	0.515(0.044)	10.225(0.459)	0.053(0.011)	26	0.519(0.044)	10.185(0.434)	0.055(0.010)
11	0.517(0.044)	10.285(0.423)	0.054(0.010)	27	0.517(0.044)	10.225(0.418)	0.051(0.010)
12	0.515(0.044)	10.218(0.419)	0.051(0.010)	28	0.518(0.044)	10.366(0.419)	0.052(0.010)
13	0.518(0.044)	10.284(0.443)	0.053(0.011)	29	0.520(0.043)	10.310(0.424)	0.052(0.011)
14	0.513(0.043)	10.405(0.485)	0.053(0.011)	30	0.515(0.043)	10.386(0.483)	0.052(0.011)
15	0.525(0.044)	10.521(0.485)	0.051(0.010)	31	0.515(0.044)	10.285(0.489)	0.053(0.011)
16	0.521(0.045)	10.193(0.418)	0.054(0.012)	32	0.515(0.044)	10.359(0.406)	0.056(0.013)

NOTE: See detailed settings of each run in Table 8 of the supplementary materials.

offloading decision based on the accurate prediction with low prediction uncertainty. Thus, the offloading decisions can be optimized via the algorithm in Chen et al. (2018) or Zhang et al. (2017) based on the predicted runtime performance metrics. While the high prediction uncertainty will prevent the adoption of the predictive offloading strategy, which highly depends on the accuracy of the predictions. Hence, other offloading strategies are preferred under this circumstance.

6. Discussion

This work develops a Bayesian regression for jointly modeling mixed multi-responses to achieve accurate prediction and uncertainty quantification with meaningful model interpretation. The proposed BS-MRMR method can quantify the hidden associations among mixed responses to improve the prediction performance. As evidenced in the case study of Fog manufacturing, the superior prediction performance of the BS-MRMR model with the capability of uncertainty quantification demonstrates its merits to support predictive offloading in Fog computing network. Not restricted to Fog manufacturing, the proposed method can also be applied in other areas such as health care and material science.

There are several directions for future researches. One direction is to investigate how to incorporate the quantified predicted uncertainty in the predictive offloading method by formulating the offloading problem as a chance-constrained optimization problem. Then the optimized offloading decisions can be more trustworthy to the performance of the predictive models. Besides predictive offloading, the proposed BS-MRMR model also facilitates the optimization of the Fog computing architecture by evaluating different designs based on the predicted performance metrics. Another direction is to extend the proposed BS-MRMR model to other types of responses such as censored outcomes and functional responses (Sun et al. 2017). For example, one may consider the proposed method for functional mixed responses, which considers runtime performance metrics as time series, hence, providing more informative prediction. Moreover, when the data are functional with no predictors available, several statistical techniques such as spline and wavelet can be applied to create a set of predictors. Then the proposed BS-MRMR method could accommodate such situations where the priors for group and individual sparsity need to be modified accordingly. Finally, we note that there are few theoretical results for the model of mixed multivariate responses due to the complex structure in the responses. For the case of single binary response and single continuous response, Kürüm et al. (2016) characterized the binary response using the latent variable and established the asymptotic normality of their estimator. It will be an interesting to borrow their ideas to investigate the posterior consistency for the proposed BS-MRMR, which is a Bayesian approach of using latent variables. Note that the proposed method adopt the spike-and-slab priors to enable variable selection, additional techniques such as Bayes factor are needed to investigate the estimation and selection consistency.

Supplementary Materials

The supplementary materials for this article contain the following: (a) detailed derivation of full-conditional distributions; (b) detailed performance comparison in numerical study; (c) detailed full factorial design for sensitivity study of prior settings; and (d) data and R implementation of the proposed BS-MRMR method for numerical study.

Acknowledgments

The authors would like to sincerely thank the editor, associate editor, and two referees for their insightful and constructive comments for helping improve the article.

Disclosure Statement

The authors report that there are no competing interests to declare.

Funding

Deng's research was supported by the National Science Foundation CISE Expedition (CCF-1918770) and Virginia Tech Data Science Faculty Fellowship. Kang's research was supported by the Natural Science Foundation of Liaoning Province (2022-MS-179).

ORCID

Xiaoyu Chen https://orcid.org/0000-0002-1870-5290
Xiaoning Kang https://orcid.org/0000-0003-0394-6240
Xinwei Deng https://orcid.org/0000-0002-1560-2405

References

Bradley, J. R. (2022), "Joint Bayesian Analysis of Multiple Response-Types Using the Hierarchical Generalized Transformation Model," *Bayesian Analysis*, 17, 127–164. [2,5,6,9,11]

Chen, X., and Jin, R. (2020), "Adapipe: A Recommender System for Aadaptive Computation Pipelines in Cyber-Manufacturing Computation Services," *IEEE Transactions on Industrial Informatics*, 17, 6221–6229. [10]

Chen, X., Wang, L., Wang, C., and Jin, R. (2018), "Predictive Offloading in Mobile-Fog-Cloud Enabled Cyber-Manufacturing Systems," in *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, pp. 167–172. IEEE. [1,10,12]

Cox, D. R., and Wermuth, N. (1992), "Response Models for Mixed Binary and Quantitative Variables," *Biometrika*, 79, 441–461. [2]

Deng, X., and Jin, R. (2015), "Qq Models: Joint Modeling for Quantitative and Qualitative Quality Responses in Manufacturing Systems," *Technometrics*, 57, 320–331. [2]

DeYoreo, M., and Kottas, A. (2018), "Bayesian Nonparametric Modeling for Multivariate Ordinal Regression," *Journal of Computational and Graphical Statistics*, 27, 71–84. [2]

Dunson, D. B. (2000), "Bayesian Latent Variable Models for Clustered Mixed Outcomes," *Journal of the Royal Statistical Society*, Series B, 62, 355–366. [2]

Fahrmeir, L., and Raach, A. (2007), "A Bayesian Semiparametric Latent Variable Model for Mixed Responses," *Psychometrika*, 72, 327–346. [2]

Fitzmaurice, G. M., and Laird, N. M. (1995), "Regression Models for a Bivariate Discrete and Continuous Outcome with Clustering," *Journal of the American statistical Association*, 90, 845–852. [2]

Hwang, B. S., and Pennell, M. L. (2014), "Semiparametric Bayesian Joint Modeling of a Binary and Continuous Outcome with Applications in Toxicological Risk Assessment," *Statistics in Medicine*, 33, 1162–1175. [2]

Kang, L., Kang, X., Deng, X., and Jin, R. (2018), "A Bayesian Hierarchical Model for Quantitative and Qualitative Responses," *Journal of Quality Technology*, 50, 290–308. [2]

Kang, X., Chen, X., Jin, R., Wu, H., and Deng, X. (2021), "Multivariate Regression of Mixed Responses for Evaluation of Visualization Designs," *IISE Transactions*, 53, 313–325. [2,3]

Kang, X., Kang, L., Chen, W., and Deng, X. (2022), "A Generative Approach to Modeling Data with Quantitative and Qualitative Responses," *Journal of Multivariate Analysis*, 190, 104952. [3]

Kürüm, E., Li, R., Shiffman, S., and Yao, W. (2016), "Time-Varying Coefficient Models for Joint Modeling Binary and Continuous Outcomes in Longitudinal Data," *Statistica Sinica*, 26, 979–1000. [3,13]

Li, Q., Pan, J., and Belcher, J. (2016), "Bayesian Inference for Joint Modelling of Longitudinal Continuous, Binary and Ordinal Events," *Statistical Methods in Medical Research*, 25, 2521–2540. [2]

Liquet, B., Mengersen, K., Pettitt, A., and Sutton, M. (2017), "Bayesian Variable Selection Regression of Multivariate Responses for Group Data," *Bayesian Analysis*, 12, 1039–1067. [3,4,6]

McCulloch, C. (2008), "Joint Modelling of Mixed Outcome Types Using Latent Variables," *Statistical Methods in Medical Research*, 17, 53–73. [2]

Ning, B., Jeong, S., and Ghosal, S. (2020), "Bayesian Linear Regression for Multivariate Responses under Group Sparsity," *Bernoulli*, 26, 2353–2382. [3]

Regan, M. M., and Catalano, P. J. (1999), "Likelihood Models for Clustered Binary and Continuous Out Comes: Application to Developmental Toxicology," *Biometrics*, 55, 760–768. [2]

Scheipl, F., Fahrmeir, L., and Kneib, T. (2012), "Spike-and-Slab Priors for Function Selection in Structured Additive Regression Models," *Journal of the American Statistical Association*, 107, 1518–1532. [4]

Stamey, J. D., Natanegara, F., and Seaman Jr, J. W. (2013), "Bayesian Sample Size Determination for a Clinical Trial with Correlated Continuous and Binary Outcomes," *Journal of Biopharmaceutical Statistics*, 23, 790–803. [2]

Sun, H., Rao, P. K., Kong, Z. J., Deng, X., and Jin, R. (2017), "Functional Quantitative and Qualitative Models for Quality Modeling in a Fused Deposition Modeling Process," *IEEE Transactions on Automation Science and Engineering*, 15, 393–403. [13]

Wagner, H., and Tüchler, R. (2010), "Bayesian Estimation of Random Effects Models for Multivariate Responses of Mixed Data," *Computational Statistics & Data Analysis*, 54, 1206–1218. [3]

Wang, H. (2015), "Scaling it Up: Stochastic Search Structure Learning in Graphical Models," *Bayesian Analysis*, 10, 351–377. [4]

Wu, D., Liu, S., Zhang, L., Terpenny, J., Gao, R. X., Kurfess, T., and Guzzo, J. A. (2017), "A Fog Computing-Based Framework for Process Monitoring and Prognosis in Cyber-Manufacturing," *Journal of Manufacturing Systems*, 43, 25–34. [1]

Wu, H., Deng, X., and Ramakrishnan, N. (2018), "Sparse Estimation of Multivariate Poisson Log-Normal Models from Count Data," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 11, 66–77. [2]

Yang, Y., Kang, J., Mao, K., and Zhang, J. (2007), "Regression Models for Mixed Poisson and Continuous Longitudinal Data," *Statistics in Medicine*, 26, 3782–3800. [2]

Yeung, W. Y., Whitehead, J., Reigner, B., Beyer, U., Diack, C., and Jaki, T. (2015), "Bayesian Adaptive Dose-Escalation Procedures for Binary and Continuous Responses Utilizing a Gain Function," *Pharmaceutical Statistics*, 14, 479–487. [2]

Zhang, K., Mao, Y., Leng, S., He, Y., and Zhang, Y. (2017), "Mobile-Edge Computing for Vehicular Networks: A Promising Network Paradigm with Predictive Off-Loading," *IEEE Vehicular Technology Magazine*, 12, 36–44. [12]

Zhang, Y., Niyato, D., and Wang, P. (2015), "Offloading in Mobile Cloudlet Systems with Intermittent Connectivity," *IEEE Transactions on Mobile Computing*, 14, 2516–2529. [1]

Zhang, Y., Wang, L., Chen, X., and Jin, R. (2019), "Fog Computing for Distributed Family Learning in Cyber-Manufacturing Modeling," in *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, pp. 88–93. IEEE. [1,10]

Zhou, Y., Whitehead, J., Bonvini, E., and Stevens, J. W. (2006), "Bayesian Decision Procedures for Binary and Continuous Bivariate Dose-Escalation Studies," *Pharmaceutical Statistics*, 5, 125–133. [2]