

# Multivariate regression of mixed responses for evaluation of visualization designs

Xiaoning Kang, Xiaoyu Chen, Ran Jin, Hao Wu & Xinwei Deng

To cite this article: Xiaoning Kang, Xiaoyu Chen, Ran Jin, Hao Wu & Xinwei Deng (2020): Multivariate regression of mixed responses for evaluation of visualization designs, IISE Transactions, DOI: [10.1080/24725854.2020.1755068](https://doi.org/10.1080/24725854.2020.1755068)

To link to this article: <https://doi.org/10.1080/24725854.2020.1755068>



Accepted author version posted online: 14 Apr 2020.  
Published online: 26 May 2020.



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



Article views: 19






View related articles [↗](#)



View Crossmark data [↗](#)

# Multivariate regression of mixed responses for evaluation of visualization designs

Xiaoning Kang<sup>a</sup> , Xiaoyu Chen<sup>b</sup>, Ran Jin<sup>b</sup> , Hao Wu<sup>c</sup>, and Xinwei Deng<sup>d</sup> 

<sup>a</sup>Institute of Supply Chain Analytics and International Business College, Dongbei University of Finance and Economics, Dalian, China; <sup>b</sup>Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA; <sup>c</sup>Discovery Analytics Center, Virginia Tech, Blacksburg, VA, USA; <sup>d</sup>Department of Statistics, Virginia Tech, Blacksburg, VA, USA

## ABSTRACT

Information visualization significantly enhances human perception by graphically representing complex data sets. The variety of visualization designs makes it challenging to efficiently evaluate all possible designs catering to users' preferences and characteristics. Most existing evaluation methods perform user studies to obtain multivariate qualitative responses from users via questionnaires and interviews. However, these methods cannot support online evaluation of designs, as they are often time-consuming. A statistical model is desired to predict users' preferences on visualization designs based on non-interference measurements (i.e., wearable sensor signals). In this work, we propose a Multivariate Regression of Mixed Responses (MRMR) to facilitate quantitative evaluation of visualization designs. The proposed MRMR method is able to provide accurate model prediction with meaningful variable selection. A simulation study and a user study of evaluating visualization designs with 14 effective participants are conducted to illustrate the merits of the proposed model.

## ARTICLE HISTORY

Received 7 October 2019  
Accepted 28 March 2020

## KEYWORDS

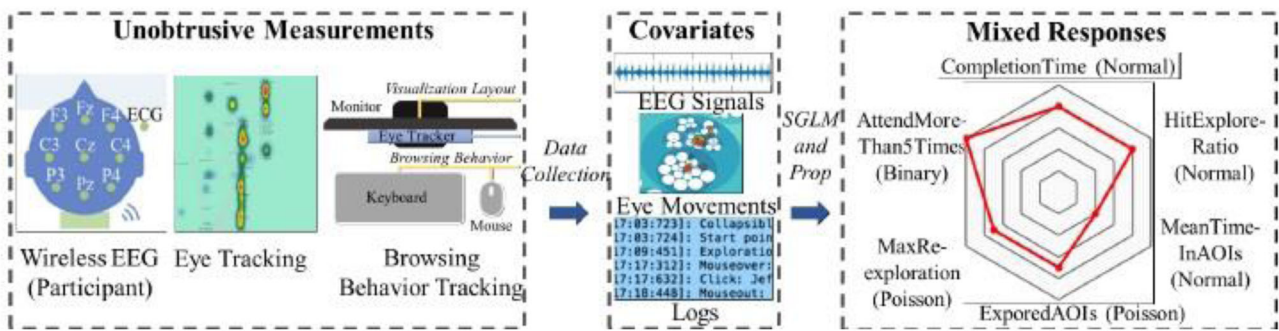
Generalized linear regression; GLASSO; information visualization; mixed responses; quantitative evaluation; wearable sensors

## 1. Introduction

Comprehending complex data sets is a cognitive challenge; however, it is an important task for human-machine collaborations. To reduce the mental workload and to facilitate fast comprehension of data, information visualization techniques have been widely applied to graphically represent complex data sets using different visualization designs. For example, visualization techniques have been adopted for visualizing manufacturing simulations (Rohrer, 2000), sequence management (Sackett *et al.*, 2006), and virtual design of factories (Lindskog *et al.*, 2013). Using a mixture of reality and virtual objects, Chen *et al.* (2016) proposed an Augmented Reality (AR)-based visualization platform to visualize real-time data streams online and incorporated an online quality-process model for a fused deposition modeling process. AR-based and virtual reality-based visualization systems have also been developed for various applications in human-machine collaboration, such as facility planning (Dangelmaier *et al.*, 2005), production control (Damiani *et al.*, 2018), and workforce training (Wang *et al.*, 2017). With a large number of visualization design candidates being available, it remains an open challenge on how to efficiently evaluate designs and select the best visualization designs catering to users' characteristics (e.g., preferences, perceptual, and cognitive capabilities) and contexts (e.g., tasks, devices, and environments).

This research concerns the online evaluation of visualization designs considering users' characteristics and contexts. In Human-Computer Interaction (HCI) as well as Human Factors and Ergonomics (HFE) communities, visualization designs have been evaluated from different perspectives by using various well-established instruments. Generally, qualitative evaluation methods have been proposed based on multivariate evaluation metrics that are collected by using questionnaires, think-aloud protocols, and interviews (Salvendy, 2012). However, these methods are typically time-consuming, due to the data collection methods. Therefore, existing qualitative methods are not very suitable to support online evaluation of visualization designs. Most of these methods also fail to quantify individual differences in users' characteristics and contexts. Although quantitative methods have been proposed to evaluate visualization designs by collecting objective data in an unobtrusive manner, most of them either fail to consider users' status (Ivory and Hearst, 2001), or can only provide a univariate response that provides limited insights to the investigators (Chen and Jin, 2017).

Obviously, a single response may not be as informative as multivariate responses to support in-depth evaluation of visualization designs from multiple perspectives. As shown in the information visualization example in Figure 1, multivariate responses are commonly collected with mixed types (i.e., continuous, counting, and binary responses). Specifically, Figure 1 shows a evaluation user study by using unobtrusive data collection devices (i.e., wireless ElectroEncephalGram (EEG) device,



**Figure 1.** A graphical illustration of data collection methods based on a wireless electroencephalogram device, eye tracking device, and a browsing behavior logging system. Three types of covariates and six responses were extracted. (This figure is partially redrawn from Chen and Jin (2017) with authors' permission)

remote eye tracking device, and a browsing behavior logging system). After collecting the data, three types of measurements (i.e., EEG signals, eye movements, and behavioral logs) with six mixed responses were extracted, including three continuous responses (i.e., *CompletionTime*, *HitExploreRatio*, and *MeanTimeInAOIs*), two counting responses (i.e., *ExploredAOIs* and *MaxRe-exploration*), and one binary response (i.e., *AttendMoreThan5Times*) (see details in Section 6). To evaluate the visualization designs in a comprehensive manner, a multivariate regression procedure for the mixed responses is required. Such a data-driven model should be able to: (i) predict mixed responses by considering the associations between themselves; and (ii) select significant predictor variables with meaningful interpretations.

In this work, we focus on developing a new Multivariate Regression for Mixed Responses (MRMR) approach, which is to jointly model the mixed responses, and hence to improve the quantitative evaluation of the online visualization designs in terms of accuracy. The proposed method exploits the hidden associations among the multiple mixed responses, which draws a clear contrast with existing methods with a separate model for each response. Clearly, ignoring the dependency among multiple responses may lose information from data. For the proposed MRMR method, we have considered a joint multivariate generalized linear model to fit data with mixed responses. The key idea is to capture the associations between multiple responses via a graphical model and improve the prediction performance via penalized regression. It is thus able to select significant predictor variables such that the model is interpretable. The Monte Carlo EM (MCEM) algorithm is adopted for efficient parameter estimation.

The rest of this article is organized as follows. Section 2 provides a literature review. Section 3 introduces the proposed MRMR model and derives the objective function. The MCEM algorithm is developed to solve the proposed estimate in Section 4. Section 5 validates and highlights the proposed MRMR model by a numerical study. An application of a visualization evaluation user study is thoroughly discussed in Section 6. We conclude this work with some discussion in Section 7.

## 2. Literature review

In this section, we review the existing literature related to the qualitative and quantitative visualization design

evaluation methods, as well as mixed response modeling, respectively, in three subsections.

### 2.1. Qualitative visualization design evaluation

Generally, existing methods for visualization design evaluation can be categorized into two groups: qualitative and quantitative evaluation methods. Qualitative methods typically provide multivariate evaluation metrics that are subjectively rated by users and can be used for the generation of insightful conclusions for investigators. For example, testbed evaluation (Bowman *et al.*, 1999) and sequential evaluation (Gabbard *et al.*, 1999) methods were proposed for usability evaluation by using heuristics. The responses collected from user studies effectively identified usability issues against design principles from multiple perspectives (Bowman *et al.*, 2002). As another example, pluralistic walkthrough (Bias, 1994) and cognitive walkthrough (Rieman *et al.*, 1995) evaluated visualization designs by collecting users' subjective responses in a series of pre-defined tasks based on their perceptual and cognitive workload. Most of the aforementioned qualitative methods collect data based on well-designed questionnaires, think-aloud protocols, and interviews (Salvendy, 2012). It typically takes hours to collect the data and days to extract informative data from records manually, and thus, they are time-consuming when used to support visualization evaluation in a timely and online manner. In addition to these qualitative evaluation methods, various instruments such as NASA task load index (TLX) (Hart and Staveland, 1988), Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988), and software usability measurement inventory (Kirakowski and Corbett, 1993) have been widely adopted in collecting the subjective feedback from user studies to provide multivariate evaluation scores. Although these methods and instruments can support in-depth evaluation of visualization designs, they need a large amount of time for user studies and response collection, and they bring interference to the HCIs.

### 2.2. Quantitative visualization design evaluation

On the other hand, quantitative methods have been reported to evaluate visualization designs by collecting objective data in an unobtrusive manner. For example, controlled experiments such as A/B testing were reported to evaluate the causal

relationship between changes and their influence on measurable behaviors (Kohavi *et al.*, 2009). An automated usability evaluating approach has been proposed by using the background logging system to record relevant interaction information about users and visualization designs (Ivory and Hearst, 2001). These quantitative methods provide objective evaluations with minimal interference, but they fail to consider the users' characteristics, which may have significant impact on the evaluation conclusions and insights. In contrast, some recent studies in HFE started to investigate data fusion and statistical models to predict subjective responses based on unobtrusive measures (e.g., physiological sensor signals, mouse and keyboard tracking, etc.). For instance, Chen and Jin (2017) employed a regularized linear regression model to accurately predict users' subjective rating of task complexities based on wearable sensor signals (i.e., EEG signals, eye movements, and browsing behaviors). Their method provided an online and quantitative evaluation model which yielded the first attempt to link unobtrusive measurements and subjective ratings. As another example, Maman *et al.*, (2017) built a regression model to predict a subjective rating of physical fatigue by using wearable sensor signals (i.e., heart rate and vibrations of human bodies). Note that the aforementioned data-driven models often consider a single response based upon wearable sensor signals (i.e., objective data) to predict subjective ratings. Therefore, it is relatively difficult for these data-driven methods to provide insights as informative as multivariate responses in qualitative methods.

### 2.3. Mixed response modeling

In the literature, there are some works on the joint modeling of multiple responses. However, most related works limited their investigation on the joint modeling for mixed responses of only continuous and binary responses (Fitzmaurice and Laird, 1997; Qiu 2008; Song *et al.*, 2009; Chen *et al.*, 2014; Deng and Jin, 2015; Kang *et al.*, 2018). They often considered a conditional model of one-type variables conditioned on the other-type variables. In addition, Hapugoda *et al.* (2017) studied a set of poultry industry data with one continuous response and one counting response. They formulated a joint model by combining a discrete time-hazard model (Hapugoda and Sooriyarachchi, 2016) and Poisson regression model. The generalized linear models were used to form marginal models for each response. Then the two responses were linked by structuring a covariance matrix to account for the potential correlations. However, such a covariance matrix contains a user-specified covariance structure. Wu *et al.* (2018) proposed to jointly model data with only multivariate counting responses by the latent variables. The R package SabreR (Berridge and Crouchley, 2011) can model data with mixed responses by using the marginal likelihood, but is limited to dealing with at most three response variables. Bonat and Jørgensen (2016) proposed a multivariate covariance generalized linear model to fit mixed responses and capture the relationship between response variables by the generalized Kronecker product. However, their method needs to pre-specify a covariance link function and several known matrices to reflect the covariance structure of

responses, and obviously such known matrices are subjectively defined by the users. Therefore, based on our best knowledge, few works have contributed to the analysis of mixed responses data with continuous, counting and binary outcomes together. A mixed response model is desired to consider the dependency among multiple responses with mixed types to support insightful online visualization evaluation.

### 3. The proposed model with exponential family

In this section, we construct the proposed model and derive its likelihood function. Suppose  $\mathbf{U} = (U^{(1)}, U^{(2)}, \dots, U^{(l)})^T$  is a random vector of  $l$  dimensions and each  $U^{(i)}$  is a normal random variable representing continuous response. Similarly,  $\mathbf{Z} = (Z^{(1)}, Z^{(2)}, \dots, Z^{(m)})^T$  is a random vector of  $m$  dimensions and each  $Z^{(i)}$  is a Poisson random variable representing counting response, and  $\mathbf{W} = (W^{(1)}, W^{(2)}, \dots, W^{(k)})^T$  is a random vector of  $k$  dimensions and each  $W^{(i)}$  is a Bernoulli random variable representing binary response. Then, we have:

$$\begin{aligned} U^{(i)} | \mu^{(i)}, \sigma^{(i)} &\sim N(\mu^{(i)}, \sigma^{(i)2}), \\ Z^{(i)} | \beta^{(i)} &\sim \text{Poisson}(\beta^{(i)}), \\ W^{(i)} | \gamma^{(i)} &\sim \text{Bernoulli}(\gamma^{(i)}), \end{aligned} \quad (1)$$

where  $\mu^{(i)}$ ,  $\sigma^{(i)2}$ ,  $\beta^{(i)}$ , and  $\gamma^{(i)}$  are the corresponding distribution parameters. Given such parameters, the model (1) implies that the variables  $U^{(i)}$ ,  $Z^{(i)}$ , and  $W^{(i)}$  are conditionally independent from each other. Let  $\xi_G = (\mu^{(1)}, \dots, \mu^{(l)})^T$ ,  $\xi_P = (\log \beta^{(1)}, \dots, \log \beta^{(m)})^T$  and

$$\xi_B = \left( \log \frac{\gamma^{(1)}}{1 - \gamma^{(1)}}, \dots, \log \frac{\gamma^{(k)}}{1 - \gamma^{(k)}} \right)^T.$$

Consequently, we connect the vector of parameters  $\xi = (\xi_G^T, \xi_P^T, \xi_B^T)^T$  with the  $p$ -dimensional vector of predictor variables  $\mathbf{x}$  via the linear model:

$$\begin{aligned} \xi &= \mathbf{B}^T \mathbf{x} + \varepsilon, \\ \varepsilon &\sim N(\mathbf{0}, \Sigma), \end{aligned}$$

where  $\mathbf{B}$  is a  $p \times (l + m + k)$  coefficient matrix, and  $\Sigma$  is the covariance matrix of the error term  $\varepsilon$ , which characterizes the dependency between multiple responses. From a Bayesian perspective, the parameter vector  $\xi$  is given a normal distribution with mean  $\mathbf{B}^T \mathbf{x}$  and covariance matrix  $\Sigma$ . Denote the parameter vector by  $\theta = (\mu^{(1)}, \dots, \mu^{(l)}, \beta^{(1)}, \dots, \beta^{(m)}, \gamma^{(1)}, \dots, \gamma^{(k)})^T$ . Accordingly, we have an element-wise mapping  $\pi$  such that  $\theta = \pi(\xi)$ . Specifically,  $\pi$  is a mapping on an  $(l + m + k)$  dimensional vector. For the first  $l$  components of the vector:  $\xi \xrightarrow{\pi} \xi$ ; for the next  $m$  components of the vector:  $\xi \xrightarrow{\pi} \exp(\xi)$ ; for the last  $k$  components of the vector:  $\xi \xrightarrow{\pi} \frac{\exp(\xi)}{1 + \exp(\xi)}$ . Hence, we propose the following multivariate regression model:

$$\begin{aligned} \mathbf{Y} &\sim \text{Exponential Family}(\theta), \\ \theta &= \pi(\mathbf{B}^T \mathbf{x} + \varepsilon), \\ \varepsilon &\sim N(\mathbf{0}, \Sigma), \end{aligned} \quad (2)$$

where  $\mathbf{Y} = [\mathbf{U}^T, \mathbf{Z}^T, \mathbf{W}^T]^T$  represents the vector of the response variables. Through the parameter vector  $\theta$ , we are

able to model the covariance structure of the variable  $Y$  indirectly.

Before diving into the derivation of the likelihood of model (2), we first examine the normal component in the regression model:

$$\begin{aligned} U &\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_U), \\ \boldsymbol{\mu} &= \mathbf{B}_U^T \mathbf{x} + \varepsilon, \\ \varepsilon &\sim N(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned} \quad (3)$$

where  $\mathbf{B}_U$  and  $\boldsymbol{\Sigma}_U$  are corresponding coefficient and covariance matrices that are related to continuous responses. Note that in this sub-model (3), the covariance structure  $\boldsymbol{\Sigma}_U$  of the responses  $U$  is coupled with that of the error term  $\varepsilon$ . To avoid the model identifiability issue, we simplify the normal sub-model as follows:

$$\begin{aligned} U &= \boldsymbol{\mu}, \\ \boldsymbol{\mu} &= \mathbf{B}_U^T \mathbf{x} + \varepsilon, \\ \varepsilon &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_U). \end{aligned}$$

In this way, the covariance structure of responses  $U$  can be modeled via the normally distributed error term  $\varepsilon$ , which reduces the number of parameters in the multivariate regression model.

With the simplification of the normal component in the regression model, its likelihood function can be derived. Under the conditional independence assumption that each dimension of  $Y$  is independent of each other given  $\boldsymbol{\theta}$ , the joint distribution of  $Y$  is

$$\begin{aligned} p(\mathbf{Y}|\boldsymbol{\theta}) &= \prod_{i=1}^l p(U^{(i)} = u^{(i)}|\boldsymbol{\theta}) \cdot \prod_{i=1}^m p(Z^{(i)} = z^{(i)}|\boldsymbol{\theta}) \cdot \prod_{i=1}^k p(W^{(i)} = w^{(i)}|\boldsymbol{\theta}) \\ &= \prod_{i=1}^m \frac{(\beta^{(i)})^{z^{(i)}} \exp(-\beta^{(i)})}{z^{(i)}!} \cdot \prod_{i=1}^k (\gamma^{(i)})^{w^{(i)}} (1 - \gamma^{(i)})^{1-w^{(i)}}. \end{aligned}$$

From  $\boldsymbol{\xi} = \mathbf{B}^T \mathbf{x} + \varepsilon$  and  $\varepsilon \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ , it is easy to derive the density function of  $\boldsymbol{\xi}|\mathbf{x}$  as

$$p(\boldsymbol{\xi}|\mathbf{x}) = \frac{1}{(2\pi)^{(l+m+k)/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\xi} - \mathbf{B}^T \mathbf{x})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\xi} - \mathbf{B}^T \mathbf{x})\right).$$

As a result, the density function of  $\boldsymbol{\theta}|\mathbf{x}$  can be written as

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}[\boldsymbol{\pi}^{-1}(\boldsymbol{\theta}) - \mathbf{B}^T \mathbf{x}]^T \boldsymbol{\Sigma}^{-1} [\boldsymbol{\pi}^{-1}(\boldsymbol{\theta}) - \mathbf{B}^T \mathbf{x}]\right)}{(2\pi)^{(l+m+k)/2} |\boldsymbol{\Sigma}|^{1/2} \prod_{i=1}^m \beta^{(i)} \prod_{i=1}^k \gamma^{(i)} (1 - \gamma^{(i)})},$$

where  $\boldsymbol{\pi}^{-1}$  is the inverse of mapping  $\boldsymbol{\pi}$ . Hence, the data distribution for  $Y|\mathbf{x}$  is

$$p(\mathbf{Y}|\mathbf{x}) = \int_{\boldsymbol{\theta}} p(\mathbf{Y}, \boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} p(\mathbf{Y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}. \quad (4)$$

Given the training data  $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)$ , we write the log-likelihood of the regression model (2) as follows:

$$\mathcal{L}(\mathbf{B}, \boldsymbol{\Sigma}) = \sum_{j=1}^n \log p(\mathbf{y}_j|\mathbf{x}_j).$$

To jointly enforce the sparsity in the estimated model parameters, the LASSO-type regularization (Tibshirani,

1996) is imposed on the matrices  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$  simultaneously. The penalty for  $\mathbf{B}$  can recover the sparse structures on model coefficients and the penalty for  $\boldsymbol{\Sigma}^{-1}$  is to enhance model prediction by incorporating the dependent relationship among the mixed responses. Consequently, the regularized negative log-likelihood is

$$\mathcal{L}_p(\mathbf{B}, \boldsymbol{\Sigma}) = -\mathcal{L}(\mathbf{B}, \boldsymbol{\Sigma}) + \lambda_1 \|\mathbf{B}\|_1 + \lambda_2 \|\boldsymbol{\Sigma}^{-1}\|_1,$$

where  $\lambda_1 > 0$  and  $\lambda_2 > 0$  are two tuning parameters, and  $\|\cdot\|_1$  stands for the  $L_1$  matrix norm, defined as  $\|\mathbf{A}\|_1 = \sum_{i,j} |a_{ij}|$  for matrix  $\mathbf{A}$  with  $a_{ij}$  being its elements. Our proposed estimates  $\hat{\mathbf{B}}$  and  $\hat{\boldsymbol{\Sigma}}$  are the solution to the following optimization problem:

$$(\hat{\mathbf{B}}, \hat{\boldsymbol{\Sigma}}) = \arg \min_{\mathbf{B}, \boldsymbol{\Sigma}} \mathcal{L}_p(\mathbf{B}, \boldsymbol{\Sigma}). \quad (5)$$

#### 4. MCEM algorithm for parameter estimation

In this section, we apply the MCEM algorithm to handle the parameter estimation in (5). The MCEM algorithm is a modified version of the EM algorithm where the expectation in the E-step is not available in a closed form. Alternatively, we approximate the expectation in the E-step by the numerical computation through Monte Carlo simulations.

##### 4.1. MCEM algorithm

In order to obtain the proposed estimates in (5), one needs to minimize the regularized negative log-likelihood function  $\mathcal{L}_p(\mathbf{B}, \boldsymbol{\Sigma})$ . However, it is a difficult task, due to the complicated integral in (4). Therefore, we employ the EM algorithm for the parameter estimation with  $\boldsymbol{\theta}$  treated as a latent variable. The following gives an overview of applying the EM algorithm to the optimization problem (5), and the detailed procedure is provided in Sections 4.2 and 4.3.

**E-step:** At iteration  $t+1$ , the conditional distribution of  $\boldsymbol{\theta}$  given  $\mathbf{x}$ ,  $\mathbf{Y}$ ,  $\mathbf{B}^{(t)}$ , and  $\boldsymbol{\Sigma}^{(t)}$  is

$$p(\boldsymbol{\theta}|\mathbf{x}, \mathbf{Y}, \mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)}) = \frac{p(\mathbf{Y}, \boldsymbol{\theta}|\mathbf{x}, \mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)})}{p(\mathbf{Y}|\mathbf{x}, \mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)})},$$

where  $\mathbf{B}^{(t)}$  and  $\boldsymbol{\Sigma}^{(t)}$  are the estimates of  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$  at iteration  $t$  in the E-step. Then the expected log-likelihood is

$$\begin{aligned} Q(\mathbf{B}, \boldsymbol{\Sigma}|\mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)}) &= E_{\boldsymbol{\theta}|\mathbf{x}, \mathbf{Y}}[\mathcal{L}(\mathbf{B}, \boldsymbol{\Sigma})] \\ &= \sum_{j=1}^n E_{\boldsymbol{\theta}|\mathbf{x}_j, \mathbf{y}_j} \left[ \log p(\mathbf{y}_j, \boldsymbol{\theta}_j|\mathbf{x}_j, \mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)}) \right]. \end{aligned} \quad (6)$$

**M-step:** Find the estimates of  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$  at iteration  $t+1$  by solving the following optimization problem:

$$\begin{aligned} (\mathbf{B}^{(t+1)}, \boldsymbol{\Sigma}^{(t+1)}) &= \arg \min_{\mathbf{B}, \boldsymbol{\Sigma}} \{-Q(\mathbf{B}, \boldsymbol{\Sigma}|\mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)}) + \lambda_1 \|\mathbf{B}\|_1 \\ &\quad + \lambda_2 \|\boldsymbol{\Sigma}^{-1}\|_1\} \end{aligned}$$

The E-step and M-step are repeated until the estimates of both  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$  are converged.

#### 4.2. E-step

In this step, we take the expectation of the log-likelihood function as in (6). However, it is difficult to derive its analytical form due to the integral in (4). Hence, the Markov Chain Monte Carlo (MCMC) technique is employed in the E-step to approximate the expected log-likelihood  $Q(\mathbf{B}, \boldsymbol{\Sigma}|\mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)})$  in (6). That is, the MCMC samples of  $\theta_j$  are drawn from:

$$\begin{aligned} p(\theta_j|\mathbf{x}_j, \mathbf{y}_j, \mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)}) \\ \propto p(\mathbf{y}_j, \theta_j|\mathbf{x}_j, \mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)}) \\ \propto \exp\left(-\frac{1}{2}[\pi^{-1}(\theta) - \mathbf{B}^T \mathbf{x}]^T \boldsymbol{\Sigma}^{-1} [\pi^{-1}(\theta) - \mathbf{B}^T \mathbf{x}]\right) \\ \exp\left(-\sum_{i=1}^m \beta^{(i)}\right) \prod_{i=1}^m [\beta^{(i)}]^{(z^{(i)}-1)} \prod_{i=1}^k [\gamma^{(i)}]^{(w^{(i)}-1)} [1 - \gamma^{(i)}]^{-w^{(i)}}. \end{aligned} \quad (7)$$

Then the expected log-likelihood function (6) is approximately computed by

$$\tilde{Q}(\mathbf{B}, \boldsymbol{\Sigma}|\mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)}) = \sum_{j=1}^n \frac{1}{h} \sum_{\nu=1}^h \log p(\mathbf{y}_j, \theta_j^{(\nu)}|\mathbf{x}_j, \mathbf{B}^{(t)}, \boldsymbol{\Sigma}^{(t)}), \quad (8)$$

where  $h$  is the size of MCMC samples after the burn-in period. In our implementation, we set the length of the MCMC chain to be 1000 with a burn-in size of 300. The MCMC chain usually converges in one or two hundred iterations.

#### 4.3. M-step

In this step, we seek for the estimates of  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$  to minimize the expected negative log-likelihood function (up to some constant):

$$\frac{1}{n} \sum_{j=1}^n \frac{1}{h} \sum_{\nu=1}^h \left\{ [\pi^{-1}(\theta_j^{(\nu)}) - \mathbf{B}^T \mathbf{x}_j]^T \boldsymbol{\Omega} [\pi^{-1}(\theta_j^{(\nu)}) - \mathbf{B}^T \mathbf{x}_j] - \log |\boldsymbol{\Omega}| \right\},$$

where  $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ . That is, we solve the following optimization problem:

$$\begin{aligned} (\mathbf{B}^{(t+1)}, \boldsymbol{\Omega}^{(t+1)}) = \arg \min_{\mathbf{B}, \boldsymbol{\Omega}} \left\{ \frac{1}{nh} \text{tr}(\boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{\Omega}) - \log |\boldsymbol{\Omega}| \right. \\ \left. + \lambda_1 \|\mathbf{B}\|_1 + \lambda_2 \|\boldsymbol{\Omega}\|_1 \right\}, \end{aligned} \quad (9)$$

where

$$\boldsymbol{\Phi} = \begin{pmatrix} \Psi(\boldsymbol{\Theta}_1) - \mathbb{X}_1 \mathbf{B} \\ \Psi(\boldsymbol{\Theta}_2) - \mathbb{X}_2 \mathbf{B} \\ \vdots \\ \Psi(\boldsymbol{\Theta}_n) - \mathbb{X}_n \mathbf{B} \end{pmatrix},$$

$\boldsymbol{\Theta}_j = (\theta_j^{(1)}, \theta_j^{(2)}, \dots, \theta_j^{(h)})^T$  is an  $(nh) \times (l + m + k)$  matrix containing all the samples of  $\theta_j$  drawn from the MCMC in the E-step. Here  $\Psi(\boldsymbol{\Theta}_j) = [\pi^{-1}(\theta_j^{(1)}), \pi^{-1}(\theta_j^{(2)}), \dots, \pi^{-1}(\theta_j^{(h)})]^T$ , and  $\mathbb{X}_j$  is an  $h \times p$  matrix with each of its rows being  $\mathbf{x}_j$ . Notice that the optimization problem (9) has two parameters  $\mathbf{B}$  and  $\boldsymbol{\Omega}$ , and it is convex with respect to one parameter when the

other is fixed. Hence, the profile technique is used to solve the optimization problem (9). More precisely, for a known estimate  $\mathbf{B}_0$ :

$$\boldsymbol{\Omega}(\mathbf{B}_0) = \arg \min_{\boldsymbol{\Omega}} \left\{ \frac{1}{nh} \text{tr}(\boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{\Omega}) - \log |\boldsymbol{\Omega}| + \lambda_2 \|\boldsymbol{\Omega}\|_1 \right\}.$$

It has the same form as Graphical LASSO (GLASSO), which has been extensively studied in the literature, for example, Yuan and Lin (2007), Friedman *et al.* (2008), Rocha *et al.* (2008), Rothman *et al.* (2008), Deng and Yuan (2009), Lam and Fan (2009), Raskutti *et al.* (2009), (2009), and Yuan (2010). On the other hand, for a known estimate  $\boldsymbol{\Omega}_0$ :

$$\mathbf{B}(\boldsymbol{\Omega}_0) = \arg \min_{\mathbf{B}} \left\{ \frac{1}{nh} \text{tr}(\boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{\Omega}_0) + \lambda_1 \|\mathbf{B}\|_1 \right\}. \quad (10)$$

To solve the optimization problem (10), we approximate the  $L_1$  matrix norm  $\|\mathbf{B}\|_1$  by a quadratic form in order to reduce the computational burden. Such a quadratic approximation technique is commonly used in the literature of sparse regressions, such as the SCAD penalty in Fan and Li (2001) and Zou and Li (2008). Denote by  $\hat{\mathbf{B}}$  the current estimate of  $\mathbf{B}$ . Let  $1/\sqrt{|\hat{\mathbf{B}}|}$  be a matrix each of whose entries is the inverse of the squared root of the absolute value for the corresponding entry in matrix  $\hat{\mathbf{B}}$ . Then we have:

$$\lambda_1 \|\mathbf{B}\|_1 \approx \lambda_1 \text{tr}(\tilde{\mathbf{B}}^T \tilde{\mathbf{B}}), \quad \text{with } \tilde{\mathbf{B}} = \mathbf{B} \circ \frac{1}{\sqrt{|\hat{\mathbf{B}}|}},$$

where  $\circ$  represents the Hadamard product (Horn and Johnson, 1985). Accordingly, the optimization problem (10) can be written as

$$\begin{aligned} \eta(\mathbf{B}) &= \frac{1}{nh} \text{tr}(\boldsymbol{\Phi}^T \boldsymbol{\Phi} \boldsymbol{\Omega}_0) + \lambda_1 \text{tr}(\tilde{\mathbf{B}}^T \tilde{\mathbf{B}}) \\ &= \frac{1}{nh} \sum_{j=1}^n \text{tr} \left\{ [\Psi(\boldsymbol{\Theta}_j) - \mathbb{X}_j \mathbf{B}]^T [\Psi(\boldsymbol{\Theta}_j) - \mathbb{X}_j \mathbf{B}] \boldsymbol{\Omega}_0 \right\} + \lambda_1 \text{tr}(\tilde{\mathbf{B}}^T \tilde{\mathbf{B}}). \end{aligned}$$

Taking the derivative of  $\eta(\mathbf{B})$  with respect to  $\mathbf{B}$  and setting it to zero, we have:

$$\begin{aligned} \frac{2}{nh} \sum_{j=1}^n \left[ \mathbb{X}_j^T \mathbb{X}_j \mathbf{B} \boldsymbol{\Omega}_0 - \mathbb{X}_j^T \Psi(\boldsymbol{\Theta}_j) \boldsymbol{\Omega}_0 \right] + \frac{2\lambda_1}{|\hat{\mathbf{B}}|} \circ \mathbf{B} = 0 \\ \times \left( \sum_{j=1}^n \mathbb{X}_j^T \mathbb{X}_j \right) \mathbf{B} \boldsymbol{\Omega}_0 + \frac{\lambda_1 nh}{|\hat{\mathbf{B}}|} \circ \mathbf{B} = \left( \sum_{j=1}^n \mathbb{X}_j^T \Psi(\boldsymbol{\Theta}_j) \right) \boldsymbol{\Omega}_0. \end{aligned} \quad (11)$$

Applying the matrix vectorization operator  $\text{vec}(\cdot)$  for both sides of (11) yields:

$$\begin{aligned} \boldsymbol{\Omega}_0 \otimes \left( \sum_{j=1}^n \mathbb{X}_j^T \mathbb{X}_j \right) \text{vec}(\mathbf{B}) + \text{vec} \left( \frac{\lambda_1 nh}{|\hat{\mathbf{B}}|} \right) \circ \text{vec}(\mathbf{B}) \\ = \text{vec} \left( \left[ \sum_{j=1}^n \mathbb{X}_j^T \Psi(\boldsymbol{\Theta}_j) \right] \boldsymbol{\Omega}_0 \right), \end{aligned}$$

where  $\otimes$  represents the Kronecker product. As a result, we obtain the solution to the optimization problem (10):

$$\text{vec}(\mathbf{B}) = \left[ \boldsymbol{\Omega}_0 \otimes \left( \sum_{j=1}^n \mathbb{X}_j^T \mathbb{X}_j \right) + \text{diag} \left( \text{vec} \left( \frac{\lambda_1 n h}{|\hat{\mathbf{B}}|} \right) \right) \right]^{-1} \\ \times \text{vec} \left( \left[ \sum_{j=1}^n \mathbb{X}_j^T \boldsymbol{\Psi}(\boldsymbol{\Theta}_j) \right] \boldsymbol{\Omega}_0 \right).$$

Based on the above discussion, the estimation procedure for the proposed model is thus summarized in Algorithm 1 as follows:

**Algorithm 1.**

**Step 1:** Input initial values  $\boldsymbol{\Sigma}_{\text{init}}, \mathbf{B}_{\text{init}}, \lambda_1$  and  $\lambda_2$ .

**Step 2:** Use the Metropolis-Hasting algorithm to draw samples of  $\boldsymbol{\theta}_j$  from (7).

$$\text{Step 3: } \boldsymbol{\Phi}^{(t+1)} = \begin{pmatrix} \boldsymbol{\Psi}(\boldsymbol{\Theta}_1) - \mathbb{X}_1 \mathbf{B}^{(t)} \\ \boldsymbol{\Psi}(\boldsymbol{\Theta}_2) - \mathbb{X}_2 \mathbf{B}^{(t)} \\ \vdots \\ \boldsymbol{\Psi}(\boldsymbol{\Theta}_n) - \mathbb{X}_n \mathbf{B}^{(t)} \end{pmatrix}.$$

**Step 4:** Obtain  $\boldsymbol{\Omega}^{(t+1)} = \text{Graphical LASSO}(\boldsymbol{\Phi}^{(t+1)}, \lambda_2)$ .

**Step 5:** Obtain  $\mathbf{B}^{(t+1)}$  as

$$\text{vec}(\mathbf{B}^{(t+1)}) = \left[ \boldsymbol{\Omega}^{(t+1)} \otimes \left( \sum_{j=1}^n \mathbb{X}_j^T \mathbb{X}_j \right) + \text{diag} \left( \text{vec} \left( \frac{\lambda_1 n h}{|\mathbf{B}^{(t)}|} \right) \right) \right]^{-1} \\ \times \text{vec} \left( \left[ \sum_{j=1}^n \mathbb{X}_j^T \boldsymbol{\Psi}(\boldsymbol{\Theta}_j) \right] \boldsymbol{\Omega}^{(t+1)} \right).$$

**Step 6:** Repeat Steps 2 - 5 till convergence.

To choose the optimal values for the tuning parameters  $\lambda_1$  and  $\lambda_2$ , we use the Extended Bayesian Information Criterion (EBIC) (Chen and Chen, 2008) which is to balance the tradeoff between the fitting of the likelihood function and the sparsity of the estimates. The EBIC is commonly used in Gaussian graphical models with sparsity when dimensions are large. Denote by  $\hat{\mathbf{B}}_{\lambda_1, \lambda_2}$  and  $\hat{\boldsymbol{\Omega}}_{\lambda_1, \lambda_2}$  the estimates of  $\mathbf{B}$  and  $\boldsymbol{\Omega}$  under the tuning parameters  $(\lambda_1, \lambda_2)$ . Then the EBIC is computed as

$$\text{EBIC}(\lambda_1, \lambda_2) = -2\tilde{Q}(\hat{\mathbf{B}}_{\lambda_1, \lambda_2}, \hat{\boldsymbol{\Omega}}_{\lambda_1, \lambda_2}) \\ + \left[ v(\hat{\mathbf{B}}_{\lambda_1, \lambda_2}) + v(\hat{\boldsymbol{\Omega}}_{\lambda_1, \lambda_2}) \right] \log n \\ + 2\tau v(\hat{\mathbf{B}}_{\lambda_1, \lambda_2}) \log [p(l+m+k)] \\ + 4\tau v(\hat{\boldsymbol{\Omega}}_{\lambda_1, \lambda_2}) \log (l+m+k),$$

where  $\tilde{Q}(\hat{\mathbf{B}}_{\lambda_1, \lambda_2}, \hat{\boldsymbol{\Omega}}_{\lambda_1, \lambda_2})$  can be approximated by (8). Here  $v(\hat{\mathbf{B}}_{\lambda_1, \lambda_2})$  and  $v(\hat{\boldsymbol{\Omega}}_{\lambda_1, \lambda_2})$  represent the number of non-zeros in the estimates  $\hat{\mathbf{B}}_{\lambda_1, \lambda_2}$  and  $\hat{\boldsymbol{\Omega}}_{\lambda_1, \lambda_2}$ , respectively.  $\tau$  is set to be 0.5 as suggested by Foygel and Drton (2010). The optimal values for  $(\lambda_1, \lambda_2)$  are chosen to minimize the EBIC.

## 5. Numerical simulation

In this section, the performance of the proposed method is examined by comparing with the separate models for each response without considering the dependency between response variables. We use the following three structures for matrix  $\boldsymbol{\Omega}$  :

**Example 1.**  $\boldsymbol{\Omega}_1 = \mathbf{L}^T \mathbf{L}$ . The elements of matrix  $\mathbf{L}$  are randomly generated from the uniform distribution  $Unif(-1, 1)$ .

**Example 2.**  $\boldsymbol{\Omega}_2 = \text{MA}(0.8, 0.6, 0.4, 0.2)$ . The main diagonal elements are ones with the  $i$ th sub-diagonal elements  $0.2 * (5 - i), i = 1, 2, 3, 4$ .

**Example 3.**  $\boldsymbol{\Omega}_3$  is generated by randomly permuting rows and corresponding columns of  $\boldsymbol{\Omega}_2$ .

Example 1 is a dense matrix, which can be considered as the most common matrix for real-world data, such as social, financial and economic data. Example 2 has a banded sparse structure taking ones on its diagonal and decreasing values on several sub-diagonals with the remaining elements being zeros. This implies that variables that are far apart are weakly correlated. This type of matrix usually occurs in longitudinal, spatial, and time series data. Example 3 considers an unstructured sparse matrix, which is often the case for high-dimensional data, such as gene expressions and image data. For each example, two scenarios of dimensions are considered: (i)  $p = 30, l = 3, m = 3$  and  $k = 3$ ; (ii)  $p = 70, l = 5, m = 5$  and  $k = 5$ . We independently generate  $n = 50$  training observations and 30 testing observations of the predictor matrices from the multivariate normal distribution  $N_p(\mathbf{0}, \sigma_X \mathbf{I})$ . The true values of the coefficient matrix  $\mathbf{B}$  are sampled from the uniform distribution  $Unif(a_B, b_B)$ . To enforce the sparsity, a proportion of zeros, denoted by  $s_B$ , is randomly placed into each column of matrix  $\mathbf{B}$ . The observations in the response matrix are generated based on the models (1) and (2). In order to investigate the impact of the variation in the predictor data on the performance of the proposed method, we scale the matrix  $\boldsymbol{\Sigma}$  such that the largest element in  $\boldsymbol{\Sigma}$  equals different values  $\varphi = (1, 1.8, 2.6, 3.4)$ . By tuning the data generation parameters  $\sigma_X, a_B, b_B$  and  $s_B$ , we make sure that the counting observations in the response matrix are within a reasonable range.

We compare the performance of the proposed method (Proposed) with the Separate Generalized Linear Model (SGLM), which models each response separately using LASSO-penalized generalized linear regression (Koh *et al.*, 2007; Park and Hastie, 2007). Note that the SGLM ignores the associations between responses. We implement the SGLM using the *glmnet* package in the R program. The Bayesian Information Criterion (BIC) is used to choose the optimal tuning parameters for SGLM.

To evaluate the estimation accuracy of the proposed method with respect to the coefficient matrix  $\mathbf{B}$  and precision matrix  $\boldsymbol{\Omega}$ , we consider the loss measures as follows:

$$L(\hat{\mathbf{B}}) = \|\mathbf{B} - \hat{\mathbf{B}}\|_F^2 \quad \text{and} \quad L(\hat{\boldsymbol{\Omega}}) = \|\boldsymbol{\Omega} - \hat{\boldsymbol{\Omega}}\|_F^2,$$

where  $\|\cdot\|_F$  represents the Frobenius norm. Here,  $\hat{\mathbf{B}}$  and  $\hat{\boldsymbol{\Omega}}$  stand for estimates of  $\mathbf{B}$  and  $\boldsymbol{\Omega}$ , respectively. Table 1 reports the averages and corresponding standard errors of the loss measures for the estimates  $\hat{\mathbf{B}}$  and  $\hat{\boldsymbol{\Omega}}$  over 50 replicates. It is clear to see that the proposed method consistently outperforms the SGLM in all settings. As the variation in the simulated data increases ( $\varphi$  increases), the traditional

**Table 1.** The averages and standard errors (in parenthesis) of loss measures for estimates.

	$\varphi$	$L(\hat{\mathbf{B}})$				$L(\hat{\mathbf{\Omega}})$	
		$p = 30$		$p = 70$		$p = 30$	$p = 70$
		SGLM	Proposed	SGLM	Proposed	Proposed	Proposed
$\Omega_1$	1.0	9.622 (1.244)	5.800 (0.047)	11.13 (0.209)	9.561 (0.076)	11.76 (0.075)	27.31 (0.041)
	1.8	18.42 (1.636)	5.652 (0.058)	13.81 (0.289)	9.458 (0.065)	11.83 (0.105)	24.02 (0.041)
	2.6	35.40 (2.104)	5.590 (0.085)	17.62 (0.781)	9.570 (0.054)	10.58 (0.058)	21.63 (0.044)
	3.4	47.09 (1.819)	5.917 (0.055)	21.43 (0.632)	9.700 (0.070)	12.76 (0.041)	27.33 (0.173)
$\Omega_2$	1.0	10.84 (0.537)	5.736 (0.080)	11.21 (0.161)	9.508 (0.070)	5.503 (0.104)	8.012 (0.008)
	1.8	13.50 (0.924)	5.690 (0.057)	13.56 (0.162)	9.446 (0.070)	5.299 (0.093)	8.035 (0.004)
	2.6	18.67 (1.549)	5.720 (0.049)	16.56 (0.277)	9.649 (0.058)	5.195 (0.092)	8.043 (0.004)
	3.4	35.64 (2.283)	5.968 (0.059)	22.02 (0.463)	9.709 (0.075)	5.422 (0.100)	8.031 (0.007)
$\Omega_3$	1.0	9.511 (0.849)	5.823 (0.113)	11.07 (0.145)	9.397 (0.062)	5.332 (0.104)	8.061 (0.007)
	1.8	14.07 (1.285)	5.780 (0.060)	13.45 (0.160)	9.402 (0.069)	5.375 (0.109)	8.050 (0.005)
	2.6	23.50 (2.070)	5.796 (0.047)	17.62 (0.286)	9.596 (0.076)	5.447 (0.107)	8.071 (0.005)
	3.4	32.50 (2.536)	5.930 (0.094)	21.81 (0.479)	9.771 (0.065)	5.681 (0.120)	8.067 (0.008)

method SGLM produces large loss measures regarding  $L(\hat{\mathbf{B}})$ , whereas the proposed method shows a more stable performance. In addition, the proposed model is designated to be able to capture the sparse structure in the underlying matrix. Thus, it can provide more accurate estimation if the underlying matrix is sparse, which is evidenced by the results of  $L(\hat{\mathbf{\Omega}})$  that the losses for  $\Omega_2$  and  $\Omega_3$  are relatively smaller than those for  $\Omega_1$ . Such results demonstrate that the proposed MRMR model improves the estimation accuracy by modeling the dependency structure between the multivariate mixed responses.

To further investigate the performance of the proposed method, we evaluate its prediction accuracy using the Root-Mean-Square Error (RMSE) for the continuous and counting responses over the testing data. With respect to the binary response, we compare the Misclassification Errors (MEs) produced by two models on the testing data. Let  $L(\text{Norm})$  and  $L(\text{Poisson})$  denote the RMSE for the estimates of continuous responses and counting responses. Let  $L(\text{Binary})$  be the ME for the estimates of binary responses. The cut-off point for the binary response estimates is 0.5. Tables 2 and 3 display the averages and corresponding standard errors of RMSE and ME for settings  $p = 30$  and  $p = 70$  over 50 replicates. For the continuous responses, although the proposed method is slightly inferior to the SGLM when the variation in the data is small ( $\varphi = 1, 1.8$ ), it has a better performance when the variation becomes larger ( $\varphi = 2.6, 3.4$ ). Under  $L(\text{Binary})$ , the prediction performance of the proposed method is consistently superior over that of the SGLM. As for the RMSE of counting responses, the proposed method substantially outperforms the SGLM, especially when  $\varphi$  is large. These results demonstrate that the proposed MRMR model improves the prediction performance by incorporating the dependent relationship between the mixed responses. Also, the proposed method is more robust than the SGLM regarding the variation in the data. Here, we also notice that the SGLM for counting responses has numeric convergence issues sometimes in the setting of  $p = 30$ , especially when  $\varphi$  is large, which is also observed in Wu *et al.* (2018). This explains the large values of  $L(\text{Poisson})$  for SGLM in Table 2.

In the proposed model, the Metropolis–Hasting method is employed to draw samples of  $\theta_j$  from (7) in the E-step of

MCEM algorithm. Note that the length of the MCMC chain and the choice of burn-in size need to be pre-specified, whereas there is no general rule on the number of MCMC samples and burn-in period. In our numerical examples, we often observe that the MCMC chain would converge in one or two hundred iterations. Considering the computational burden, we suggest 1000 MCMC draws with the first 300 samples as the burn-in period for the proposed model. Figure 2 displays the trace plots and corresponding AutoCorrelation Function (ACF) plots for six randomly chosen parameters from one replicate of Example 1 with  $p = 30$  and  $\varphi = 1$ . The results from the figure indicate that the 1000 MCMC samples and 300 burn-in size appear to be reasonable choices. The rest of the parameters have similar patterns of trace plots and ACF plots, and hence are omitted here.

To further enhance the scope of the proposed method, we conduct a simulation setting with the comparison method of the Multivariate Covariance Generalized Linear Model (MCGLM) by Bonat and Jørgensen (2016), which is implemented using the R package *mcglm*. The MCGLM takes into account the possible association between the multivariate mixed responses when fitting them within the framework of generalized linear models. It can provide an estimate of the inverse covariance matrix characterizing the relationship between responses. However, the MCGLM needs the selection of a proper covariance link function and several known matrices before data analysis, which reflects the association structure of the responses. The choice of an appropriate covariance link function is often subjectively made based on a user's experience. Additionally, the MCGLM encounters the algorithm convergence issue for large  $p$ . A possible explanation is that their algorithm does not incorporate the regularization technique. In contrast, the proposed model takes advantage of the well-developed techniques of LASSO and GLASSO to circumvent the computational convergence for high-dimensional data. We thus run a numerical simulation on a small scale of  $p = 5$ ,  $l = 2$ ,  $m = 2$  and  $k = 1$  for Example 1. The generation of training set, testing set, the true coefficient matrix  $\mathbf{B}$ , the covariance matrix  $\mathbf{\Sigma}$ , the values of  $\varphi$  and other relevant parameters remains the same as above. Table 4 reports the averages and

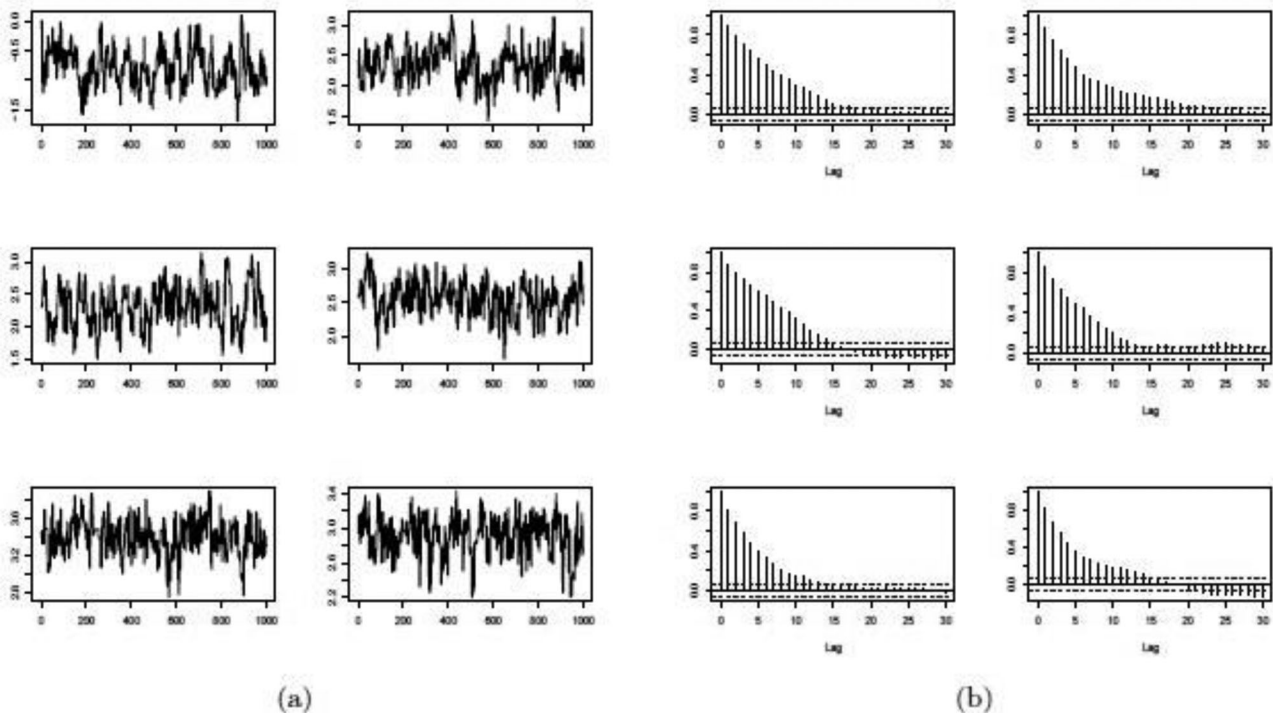


**Table 2.** The averages and standard errors (in parenthesis) of RMSE and ME when  $p = 30$ .

	$\varphi$	$L(\text{Norm})$		$L(\text{Poisson})$		$L(\text{Binary})$	
		SGLM	Proposed	SGLM	Proposed	SGLM	Proposed
$\Omega_1$	1.0	1.393 (0.020)	1.499 (0.017)	22.01 (3.275)	10.25 (1.303)	0.321 (0.004)	0.263 (0.004)
	1.8	1.664 (0.021)	1.754 (0.019)	75.21 (6.590)	11.42 (1.199)	0.325 (0.006)	0.265 (0.005)
	2.6	1.977 (0.035)	1.951 (0.024)	178.2 (37.13)	17.02 (1.634)	0.325 (0.004)	0.265 (0.005)
	3.4	2.293 (0.042)	2.111 (0.050)	633.9 (176.9)	25.10 (1.650)	0.326 (0.005)	0.266 (0.007)
$\Omega_2$	1.0	1.362 (0.018)	1.503 (0.017)	18.13 (2.943)	8.540 (0.720)	0.317 (0.006)	0.260 (0.005)
	1.8	1.640 (0.020)	1.726 (0.021)	122.1 (34.63)	14.47 (1.553)	0.322 (0.005)	0.263 (0.004)
	2.6	2.031 (0.035)	1.958 (0.025)	219.7 (61.37)	17.81 (1.499)	0.321 (0.005)	0.262 (0.004)
	3.4	2.356 (0.037)	2.142 (0.024)	844.9 (195.4)	25.20 (1.695)	0.323 (0.005)	0.270 (0.004)
$\Omega_3$	1.0	1.360 (0.019)	1.488 (0.020)	15.89 (3.713)	7.833 (0.831)	0.314 (0.004)	0.259 (0.004)
	1.8	1.674 (0.022)	1.728 (0.021)	138.9 (32.26)	12.54 (1.160)	0.302 (0.004)	0.259 (0.004)
	2.6	1.979 (0.030)	1.936 (0.021)	248.7 (52.80)	19.90 (1.537)	0.316 (0.005)	0.264 (0.004)
	3.4	2.303 (0.040)	2.132 (0.023)	779.7 (132.9)	23.94 (1.946)	0.313 (0.005)	0.267 (0.004)

**Table 3.** The averages and standard errors (in parenthesis) of RMSE and ME when  $p = 70$ .

	$\varphi$	$L(\text{Norm})$		$L(\text{Poisson})$		$L(\text{Binary})$	
		SGLM	Proposed	SGLM	Proposed	SGLM	Proposed
$\Omega_1$	1.0	1.448 (0.017)	1.511 (0.017)	9.886 (0.778)	8.270 (0.661)	0.387 (0.004)	0.343 (0.004)
	1.8	1.696 (0.019)	1.721 (0.017)	18.06 (1.740)	13.61 (1.192)	0.382 (0.005)	0.353 (0.006)
	2.6	2.133 (0.047)	1.962 (0.024)	28.30 (2.625)	19.81 (1.676)	0.395 (0.005)	0.367 (0.005)
	3.4	2.500 (0.058)	2.136 (0.036)	55.09 (3.344)	28.04 (1.740)	0.399 (0.006)	0.357 (0.004)
$\Omega_2$	1.0	1.452 (0.016)	1.520 (0.016)	11.62 (0.731)	8.026 (0.698)	0.372 (0.004)	0.333 (0.004)
	1.8	1.743 (0.019)	1.748 (0.018)	20.70 (1.584)	14.10 (0.942)	0.372 (0.006)	0.332 (0.004)
	2.6	2.074 (0.028)	1.972 (0.024)	28.99 (2.338)	19.44 (1.110)	0.366 (0.004)	0.336 (0.004)
	3.4	2.536 (0.043)	2.164 (0.022)	41.92 (3.230)	26.26 (1.595)	0.383 (0.004)	0.340 (0.003)
$\Omega_3$	1.0	1.408 (0.012)	1.467 (0.013)	13.25 (0.742)	7.565 (0.527)	0.370 (0.005)	0.330 (0.005)
	1.8	1.726 (0.016)	1.736 (0.016)	18.76 (2.338)	13.22 (1.057)	0.361 (0.004)	0.325 (0.004)
	2.6	2.154 (0.029)	1.985 (0.019)	27.42 (2.224)	18.71 (1.306)	0.379 (0.005)	0.341 (0.004)
	3.4	2.520 (0.032)	2.161 (0.016)	42.33 (3.170)	25.03 (1.228)	0.378 (0.004)	0.337 (0.004)

**Figure 2.** Trace plots and ACF plots for the randomly selected parameters of one replicate from Example 1 when  $p = 30$  and  $\varphi = 1$ . (a) trace plots; (b) ACF plots.

corresponding standard errors of all the loss measures of  $L(\hat{B})$ ,  $L(\hat{\Omega})$ ,  $L(\text{Norm})$ ,  $L(\text{Poisson})$  and  $L(\text{Binary})$  for three methods over 50 replicates. Clearly, it is seen that the proposed model is overall the best compared with the SGLM and MCGLM approaches. It dominates the other two methods in terms of  $L(\hat{B})$  and produces lower values of  $L(\hat{\Omega})$ .

Although the proposed model is comparable to the MCGLM regarding  $L(\text{Norm})$ ,  $L(\text{Poisson})$  and  $L(\text{Binary})$  when  $\varphi = 1$  and 1.8, it performs better as the value of variation parameter  $\varphi$  increases. The SGLM, which does not consider the dependency between response variables, is inferior to the MCGLM and the proposed model.

**Table 4.** The averages and standard errors (in parenthesis) of loss measures when  $p = 5$ .

		$\varphi = 1$	$\varphi = 1.8$	$\varphi = 2.6$	$\varphi = 3.4$
$L(\hat{\mathbf{B}})$	SGLM	1.052(0.061)	1.505(0.034)	1.808(0.050)	2.241(0.073)
	MCGLM	0.858(0.022)	1.372(0.031)	1.675(0.050)	2.194(0.053)
	Proposed	0.478(0.012)	0.477(0.010)	0.485(0.010)	0.509(0.011)
$L(\hat{\Omega})$	MCGLM	6.008(0.017)	4.559(0.013)	8.233(0.016)	5.660(0.117)
	Proposed	5.033(0.026)	3.411(0.014)	6.501(0.022)	4.697(0.032)
$L(\text{Norm})$	SGLM	1.051(0.018)	1.418(0.021)	1.728(0.033)	1.885(0.049)
	MCGLM	1.043(0.017)	1.397(0.021)	1.681(0.031)	1.877(0.032)
	Proposed	1.045(0.018)	1.398(0.021)	1.649(0.028)	1.826(0.029)
$L(\text{Poisson})$	SGLM	2.981(0.342)	6.202(0.449)	10.48(1.096)	19.53(2.154)
	MCGLM	2.649(0.113)	6.087(0.433)	9.779(0.948)	17.78(1.612)
	Proposed	2.666(0.121)	5.858(0.419)	8.846(0.785)	14.21(1.328)
$L(\text{Binary})$	SGLM	0.264(0.009)	0.282(0.007)	0.289(0.006)	0.292(0.008)
	MCGLM	0.243(0.007)	0.257(0.006)	0.267(0.006)	0.272(0.006)
	Proposed	0.237(0.007)	0.241(0.006)	0.243(0.006)	0.251(0.006)

## 6. Case study

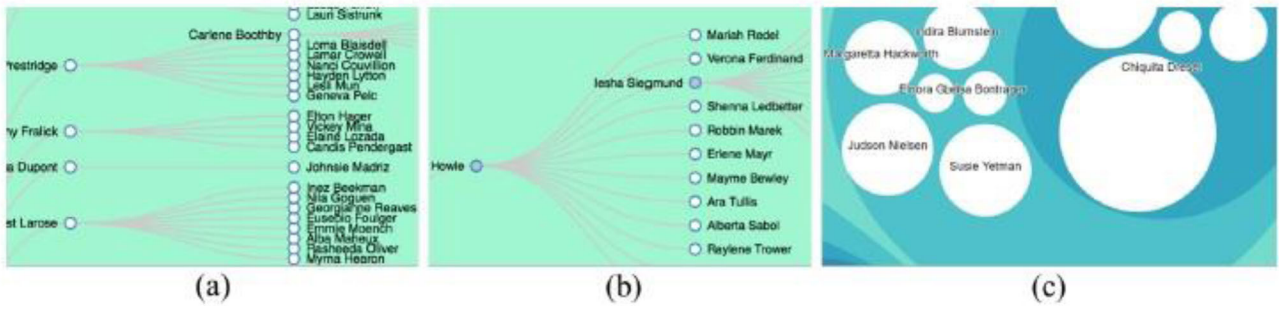
In this section, we demonstrate the merits of the proposed method through the case study of information visualization described in Section 1. This data set was collected from a user study to quantitatively evaluate three interactive visualization designs (Chen and Jin, 2017), namely, the static node-link tree diagram, the collapsible node-link tree, and the zoomable layout (see Figure 3). In this user study, 15 participants were recruited to evaluate three interactive visualization designs by performing 11 pre-defined visual searching tasks for each design. Note that data for one participant was removed due to missing values.

During performing tasks, as presented in Figure 1, all participants' EEG signals, eye movements, and browsing behaviors were collected by using an ABM<sup>®</sup> B-Alert 10-channel wireless EEG device, a SMI<sup>®</sup> REDn remote eye tracker, and a background logging system, respectively. Following the feature extraction methods described in Chen and Jin (2017), 481 predictor variables were extracted from these sensor signals and browsing behaviors. Some exploratory data analysis was conducted to screen out the unimportant predictors, which resulted in 58 predictor variables for use.

In total, the data set contains  $n = 462$  ( $14 \times 11 \times 3$ ) observations with three continuous response variables ( $Y_1, Y_2, Y_3$ ), two counting response variables ( $Y_4, Y_5$ ), and one binary response variable ( $Y_6$ ). The  $Y_1$  is *CompletionTime*, which identifies the total time consumption for one participant to perform a task. The  $Y_2$  is *HitExploreRatio*, a ratio of the number of visually reached targets over the number of explored Area Of Interests (AOI). Here AOI is referred as a node in a node-link tree diagram, which associates with eye fixation for more than 0.5 second. The  $Y_3$  is *MeanTimeInAOIs* as average time spent in AOIs, which can be used to measure the level of confusion the AOIs caused for participants. The counting response  $Y_4$  is *ExploredAOIs*, which counts the total number of explored AOIs in performing a task; and the  $Y_5$  is *MaxReexploration* as the highest number of re-explorations at one AOI in performing a task. The binary response is  $Y_6$ , *AttendMoreThan5Times*, which has value one when *MaxReexploration* is higher than five, and zero otherwise. These six response variables represent the efficiency of the visualization design from different perspectives.

We compare the performance of the proposed method with SGLM by considering two approaches to data partition for training and testing. The first approach is random splits: the whole data set is randomly partitioned into a training set with sample size 200 and a testing set with the remaining 262 observations; The second approach is leave-one-participant-out Cross-Validation (CV): one (33 observations) out of 14 participants is iteratively left out for testing, and the rest are used for training (429 observations). Note that the leave-one-participant-out CV is investigated to highlight the proposed model, since individual differences may easily lead to the violation of assumptions for a linear model in a usability test that evaluates a user interface (Nielsen and Molich, 1990). The training set is used to fit the proposed model and SGLM, and the loss measures  $L(\text{Norm})$ ,  $L(\text{Poisson})$  and  $L(\text{Binary})$  are computed from the testing data. Table 5 summarizes the averages of loss measures and corresponding standard errors from the random splits of 50 times, and the leave-one-participant-out CV.

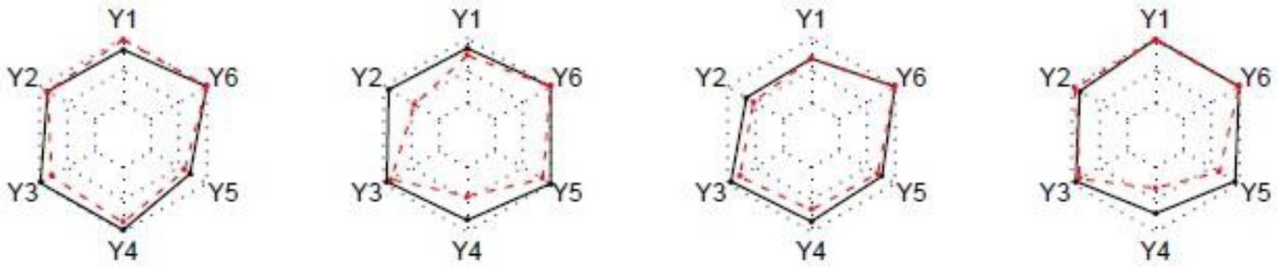
From the table, it is seen that the proposed model generally outperforms the SGLM in terms of each loss measure for both random splits and leave-one-participant-out CV. In particular, the proposed model significantly improves the prediction accuracy for the counting responses with a much smaller prediction error and standard error compared with the SGLM. The proposed model is also superior over the SGLM with respect to the continuous responses, and slightly better regarding the binary response. These results demonstrate the advantage of the proposed MRMR model over the SGLM by considering the dependency relation between multiple responses. In addition, to further evaluate the prediction ability with respect to the overall mixed responses, Figure 4 displays four radar maps corresponding to four testing data points randomly selected from one split of data. The hexagons, solid lines and dashed lines stand for the testing data values, the prediction values from the proposed MRMR and the prediction values from the SGLM, respectively. Hence, the closeness of the graphs composed of the solid or dashed lines to the hexagons illustrates the prediction ability of the corresponding models. From Figure 4, it can be seen that although the proposed MRMR model with solid lines may be inferior sometimes with certain response ( $Y_1$  for the first panel, and  $Y_2$  for the fourth panel), it generally performs better than the SGLM. Furthermore, another



**Figure 3.** Example of three visualization designs for user study, where (a) is the static node-link tree to map names and hierarchical relationships to the circles, texts, and edges, respectively; (b) is the collapsible node-link tree, which is an interactive variant of (a) with collapsible branches; and (c) is the zoomable pack layout to map the names and hierarchical relationships to nested and zoomable circles (Redrawn from Chen and Jin (2017) with authors' permission).

**Table 5.** The averages and standard errors (in parenthesis) of loss measures from 50 random splits of data and leave-one-participant-out.

		$L(Norm)$	$L(Poisson)$	$L(Binary)$
Random Splits	SGLM	3.001 (0.015)	115.3 (21.89)	0.189 (0.003)
	Proposed	2.808 (0.019)	36.72 (0.311)	0.183 (0.004)
Leave-one-participant-out	SGLM	2.966 (0.121)	47.82 (11.97)	0.186 (0.014)
	Proposed	2.769 (0.146)	36.38 (2.098)	0.178 (0.016)



**Figure 4.** Radar maps for the predictions of the proposed MRMR and SGLM from the randomly selected four samples of testing data. The hexagons represent the testing samples; the black solid lines are the predictions of the proposed MRMR; the red dashed lines indicate the predictions of the SGLM.

**Table 6.** The estimated correlation matrix of responses obtained by the proposed model from one random split of data.

	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$	$Y_6$
$Y_1$	1					
$Y_2$	-0.193	1				
$Y_3$	0.060	0.143	1			
$Y_4$	0.036	-0.011	-0.046	1		
$Y_5$	0.252	-0.085	0.009	0.256	1	
$Y_6$	0.670	-0.224	0.027	0.119	0.397	1

advantage is that the proposed method is able to provide an estimate of the correlation matrix of response variables, which captures the dependency relationship between responses. Table 6 reports such an estimate from one random split of data. Although there are some variables with weak correlations, several relatively strong correlations exist in variables  $Y_1$  and  $Y_6$ ,  $Y_5$  and  $Y_6$ ,  $Y_4$  and  $Y_5$ ,  $Y_1$  and  $Y_5$ , which cannot be ignored. We also check each estimate of  $\Omega$  obtained from 50 splits of data and find that most estimates are dense matrices. The mean of the largest absolute value in  $\Sigma$  is 1.72 with its standard deviation being 0.044. Hence, the underlying matrix  $\Omega$  of the data might be closely represented by the Example 1 with variation  $\varphi = 1.8$  in Section 5.

In addition to the statistical performance, several interesting findings from a HCI perspective can be identified from the superior performance of the proposed MRMR model.

First, the proposed model provides an accurate prediction for subjective responses in a visualization evaluation user study simply based on an objective and unobtrusive measurements and a logging system. In the literature, interviews and questionnaires are typically adopted to provide subjective responses regarding the effectiveness and efficiency of the visualization designs (Bowman *et al.*, 2002). However, they typically lead to interference in participants' subjective responses and require a huge amount of time to conduct the interviews and questionnaires. As an enhancement to Chen and Jin (2017), who were the first to prove the feasibility for quantitative and unobtrusive evaluation, the proposed MRMR method enables the joint modeling of multiple mixed responses, providing more realistic techniques to support efficient evaluation studies in HCI.

Second, the results from the leave-one-participant-out CV analysis indicate that the proposed model can better address the individual differences in the user study, as it provides a better prediction performance for all three types of responses. Such an advantage can be explained by the quantification of correlation among the responses. In a user interface evaluation study, the investigators typically design a few highly correlated questions for one evaluation criterion to reduce the effects of individual differences and randomness in participants' responses, e.g., NASA TLX (Hart and Staveland, 1988), SWAT (Reid and Nygren, 1988), etc. The

proposed MRMR model benefits from the quantification of the designed correlations among responses. Therefore, it can serve as an analytical tool for HCI researchers and practitioners to investigate the individual differences.

On the other hand, the proposed MRMR method also provides good model interpretation due to the variable selection. Based on the 95% confidence intervals constructed from 50 estimates obtained from 50 random splits of data, we found that the selected variables for  $Y_4$  include “the average length of frequent eye movement trajectories”, “the alpha and gamma band of EEG signals”, “variance of mouse moving over/out duration”, etc. The variable “average length of frequent eye movement trajectories” can reflect participants’ different scanning paths. Specifically, a higher value indicates continuous and fluent visual searching, and a lower value indicates that the participants may be confused with the visualization layout. The selection of EEG-related variables aligns well with some research findings in the literature. For example, the variable “alpha and gamma band of EEG signals” was reported to be significantly correlated with sustained attention during performing the visual searching tasks (Huang *et al.*, 2007). Therefore, the estimated coefficients of these selected variables can help to understand how the visualization designs help participants concentrate on the contents. In addition, the variable “variance of mouse moving over/out durations” reflects the participants’ visual search strategy of using the mouse to direct their attention, which is intuitively correlated with their total numbers of explored AOIs in different visualization designs. These insightful understandings generated by the variable selection results of the proposed MRMR model can further improve the visualization designs.

## 7. Discussion

In this article, we study the online evaluation problem for interactive information visualization designs based on non-invasive measurements (i.e., wearable sensor signals). This research aims at evaluating designs online catering to users’ preferences and characteristics with mixed subjective responses by using non-interference measurements. In this case, a joint model is able to improve the estimation and prediction results compared with modeling the responses separately when responses are correlated. Therefore, a generalized linear model with mixed (i.e., continuous, counting, and binary) responses is proposed to quantify the correlation between mixed responses by using a graphical model. The penalty terms are imposed on the negative likelihood function to encourage sparsity in the estimated coefficients and precision matrices, such that the model is interpretable. Both the numerical study and a user study in visualization evaluation show the advantages of the proposed MRMR method over the separate models.

The proposed MRMR model can be readily applied to the design and evaluation of information visualization in smart manufacturing, which can help novices understand complex process data sets and help experts generate insights from visualized data streams. For example, the MRMR

method can be directly applied to efficiently evaluate and improve the AR-based visualization platform proposed in Chen *et al.* (2016). Specifically, when manufacturing users (e.g., operators, engineers, and managers) are performing tasks by exploring data in an AR visualization platform, their cognitive status can be quantified by the MRMR model and thus supporting online adaptive visualization. Furthermore, the applications of the MRMR model are not limited to the evaluation of visualization designs. It can also be easily extended to support evaluation studies in HFE and HCI when multiple subjective responses are collected for in-depth understanding of human behaviors. In addition, by taking advantage of the model structure of generalized linear models, the proposed method is able to accommodate other common types of response variables with exponential-family distributions. For example, the negative inverse link function can be used if the responses follow exponential and Gamma distributions. The logit link function is used if the responses are from a Binomial distribution. The derivation of the likelihood function and the objective in Section 3 can be obtained in a similar fashion. However, the proposed modeling framework cannot cover all distributions from the exponential family, such as the Chi-square distribution and Wishart distribution. For such distributions, it is not straightforward to construct the link functions used in the generalized linear models.

There are a few directions for future work. First, to better understand and optimize the human decision making process in interacting with visualization designs, an online learning method for multi-step and multi-scale decision making will be studied. Second, when some response variables inherently have very high dimensions compared with other responses, one can seek for the conditional models which are constructed by fitting high-dimensional responses conditional on other responses. Such a conditional modeling approach would properly reflect the importance of the high-dimensional responses. The techniques of modified Cholesky decomposition (Kang *et al.*, 2019; Kang and Deng, 2020) can be potentially useful for dealing with the conditional modeling for high-dimensional responses. Third, one can extend the proposed MRMR model to accommodate the missing data, of which the parameter estimation can still be addressed under the EM framework. In particular, when different response variables have different sampling intervals, the sampling bias would potentially cause inefficiency and inaccuracy in parameter estimation. To address such an issue, one solution is to treat the unsampled data points as the missing values. Then the proposed method can be extended under the framework of missing-data imputation where the EM algorithm is commonly used for dealing with missing data. Another possible solution is to borrow strength from the Seemingly Unrelated Regressions (SUR) (Zellner, 1962) to allow different sampling intervals for different responses. However, the conventional SUR can only deal with multiple continuous responses. It will be an interesting future research topic to integrate the proposed modeling technique with seemingly unrelated regressions for the mixed responses data.

## Acknowledgments

The authors thank the Editor and referees for their insightful and helpful comments that have greatly improved the original manuscript.

## Funding

The authors would like to acknowledge the support from Ministry of Education of China (20YJC910007), and National Natural Science Foundation of China (71871047).

## Notes on contributors

**Xiaoning Kang** received his PhD degree in statistics from Virginia Tech. He is now an associate professor at the Institute of Supply Chain Analytics, and International Business College in Dongbei University of Finance and Economics, China. His research interests include high-dimensional matrix estimation, mixed responses data analysis, Bayesian hierarchical modeling and computer experiments design.

**Xiaoyu Chen** is a PhD student in the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received his Bachelor's degree in optical information science and technology from Beijing Institute of Technology in 2011. His research interests include human-machine collaboration, recommendation for both adaptive computation services and adaptive information visualization systems, and immersive data analytics.

**Ran Jin** is an associate professor at the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received his Bachelor's degree in electronic engineering from Tsinghua University in 2005, PhD degree in industrial engineering from Georgia Tech in 2011, his Master's degrees in industrial engineering (2007) and in statistics (2009), both from the University of Michigan. His research interests focus on system informatics and control in industrial engineering, quality engineering in manufacturing scale-up, and sensing, modeling and process optimization based on high definition profile data.

**Hao Wu** received his Bachelor's degree in software engineering from Tsinghua University in Beijing, China, and PhD degree in computer engineering from Virginia Tech, where he did research on maximum entropy models, multivariate regression with covariance estimation, data analytics, and machine learning. He is now a software engineer at Google Research. His research interests focus on machine learning and deep learning, especially on learning with multimodal approaches, neural network architecture search, and neural network model compression.

**Xinwei Deng** is an associate professor in the Department of Statistics at Virginia Tech. He received his Bachelor's degree in mathematics from Nanjing University, and PhD degree in industrial engineering from Georgia Tech. His research interests focus on statistical modeling and data analysis, including high-dimensional classification, graphical model estimation, and the interface between experimental design and machine learning. He is an elected member of ISI, and a member of INFORMS and ASA.

## ORCID

Xiaoning Kang  <http://orcid.org/0000-0003-0394-6240>

Ran Jin  <http://orcid.org/0000-0003-3847-4538>

Xinwei Deng  <http://orcid.org/0000-0002-1560-2405>

## References

Berridge, D.M. and Crouchley, R. (2011) *Multivariate Generalized Linear Mixed Models Using R*. CRC Press, New York, NY.

- Bias, R.G. (1994) The pluralistic usability walkthrough: Coordinated empathies, in *Usability Inspection Methods*, pp. 63–76. John Wiley & Sons, New York, NY.
- Bonat, W.H. and Jørgensen, B. (2016) Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **65**(5), 649–675.
- Bowman, D., Johnson, D. and Hodges, L. (1999) Testbed environment of virtual environment interaction, in *Proceedings of the ACM VRST'99 Symposium on Virtual Reality Software and Technologies*, pp. 26–33. Association for Computing Machinery, New York, NY.
- Bowman, D.A., Gabbard, J.L. and Hix, D. (2002) A survey of usability evaluation in virtual environments: Classification and comparison of methods. *Presence: Teleoperators and Virtual Environments*, **11**(4), 404–424.
- Chen, J. and Chen, Z. (2008) Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, **95**(3), 759–771.
- Chen, S., Witten, D.M. and Shojaie, A. (2014) Selection and estimation for mixed graphical models. *Biometrika*, **102**(1), 47–64.
- Chen, X. and Jin, R. (2017) Statistical modeling for visualization evaluation through data fusion. *Applied Ergonomics*, **65**, 551–561.
- Chen, X., Sun, H. and Jin, R. (2016) Variation analysis and visualization of manufacturing processes via augmented reality, in *Proceedings of IISE Annual Conference*.
- Damiani, L., Demartini, M., Guizzi, G., Revetria, R. and Tonelli, F. (2018) Augmented and virtual reality applications in industrial systems: A qualitative review towards the industry 4.0 era. *IFAC-PapersOnLine*, **51**(11), 624–630.
- Dangelmaier, W., Fischer, M., Gausemeier, J., Grafe, M., Matysczok, C. and Mueck, B. (2005) Virtual and augmented reality support for discrete manufacturing system simulation. *Computers in Industry*, **56**(4), 371–383.
- Deng, X. and Jin, R. (2015) QQ models: Joint modeling for quantitative and qualitative quality responses in manufacturing systems. *Technometrics*, **57**(3), 320–331.
- Deng, X. and Yuan, M. (2009) Large Gaussian covariance matrix estimation with Markov structures. *Journal of Computational and Graphical Statistics*, **18**(3), 640–657.
- Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**(456), 1348–1360.
- Fitzmaurice, G.M. and Laird, N.M. (1997) Regression models for mixed discrete and continuous responses with potentially missing values. *Biometrics*, **53**(1), 110–122.
- Foygel, R. and Drton, M. (2010) Extended Bayesian information criteria for Gaussian graphical models, in *Advances in Neural Information Processing Systems*, pp. 604–612. MIT Press, Cambridge, MA.
- Friedman, J., Hastie, T. and Tibshirani, R. (2008) Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics*, **9**(3), 432–441.
- Gabbard, J.L., Hix, D. and Swan, J.E. (1999) User-centered design and evaluation of virtual environments. *IEEE Computer Graphics and Applications*, **19**(6), 51–59.
- Hapugoda, J., Kalupahana, R., Sooriyachchi, M. and Satharasinghe, D. (2017) Joint modeling of mixed responses: An application to poultry industry, *International Conference on Computational Mathematics, Computational Geometry and Statistics*, Singapore, pp. 182–185.
- Hapugoda, J. and Sooriyachchi, M. (2016) Joint modeling of survival and counts: Joining the discrete time hazard model with Poisson regression model, in *Proceedings of the Conference of International Society of Clinical Biostatistics*, United Kingdom.
- Hart, S.G. and Staveland, L.E. (1988) Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, **52**, 139–183.
- Horn, R.A. and Johnson, C.R. (1985) *Matrix Analysis*, Cambridge University Press, Cambridge, United Kingdom.
- Huang, R.-S., Jung, T.-P. and Makeig, S. (2007) Multi-scale EEG brain dynamics during sustained attention tasks, in *Proceedings of the*

- 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, volume 4, IEEE Press, Piscataway, NJ, pp. IV-1173
- Ivory, M.Y. and Hearst, M.A. (2001) The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, **33**(4), 470–516.
- Kang, L., Kang, X., Deng, X. and Jin, R. (2018) A Bayesian hierarchical model for quantitative and qualitative responses. *Journal of Quality Technology*, **50**(3), 290–308.
- Kang, X. and Deng, X. (2020) An improved modified Cholesky decomposition approach for precision matrix estimation. *Journal of Statistical Computation and Simulation*, **90**(3), 443–464.
- Kang, X., Deng, X., Tsui, K. and Pourahmadi, M. (2019) On variable ordination of modified Cholesky decomposition for estimating time-varying covariance matrices. *International Statistical Review*, in press.
- Kirakowski, J. and Corbett, M. (1993) SMI: The software usability measurement inventory. *British Journal of Educational Technology*, **24**(3), 210–212.
- Koh, K., Kim, S.-J. and Boyd, S. (2007) An interior-point method for large-scale L1-regularized logistic regression. *Journal of Machine Learning Research*, **8**, 1519–1555.
- Kohavi, R., Longbotham, R., Sommerfield, D. and Henne, R.M. (2009) Controlled experiments on the web: Survey and practical guide. *Data Mining and Knowledge Discovery*, **18**(1), 140–181.
- Lam, C. and Fan, J. (2009) Sparsistency and rates of convergence in large covariance matrix estimation. *Annals of Statistics*, **37**(6B), 4254–4278.
- Lindskog, E., Berglund, J., Vallhagen, J. and Johansson, B. (2013) Visualization support for virtual redesign of manufacturing systems. *Procedia CIRP*, **7**, 419–424.
- Maman, Z.S., Yazdi, M.A.A., Cavuoto, L.A. and Megahed, F.M. (2017) A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Applied Ergonomics*, **65**, 515–529.
- Nielsen, J. and Molich, R. (1990) Heuristic evaluation of user interfaces, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, pp. 249–256.
- Park, M.Y. and Hastie, T. (2007) L1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, **69**(4), 659–677.
- Qiu, P. (2008) Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions*, **40**(7), 664–677.
- Raskutti, G., Yu, B., Wainwright, M.J. and Ravikumar, P.K. (2009) Model selection in Gaussian graphical models: High-dimensional consistency of L1-regularized MLE, in *Advances in Neural Information Processing Systems*. pp. 1329–1336. MIT Press, Cambridge, MA.
- Reid, G.B. and Nygren, T.E. (1988) The subjective workload assessment technique: A scaling procedure for measuring mental workload, in *Advances in Psychology*, pp. 185–218. North Holland Publishing Co., Amsterdam, Netherlands.
- Rieman, J., Franzke, M. and Redmiles, D. (1995) Usability evaluation with the cognitive walkthrough, in *Conference Companion on Human Factors in Computing Systems*, pp. 387–388. Association for Computing Machinery, New York, NY.
- Rocha, G.V., Zhao, P. and Yu, B. (2008) A path following algorithm for sparse pseudo-likelihood inverse covariance estimation (splice). *arXiv preprint arXiv:0807.3734*.
- Rohrer, M.W. (2000) Seeing is believing: The importance of visualization in manufacturing simulation, in *Proceedings of the 32nd Conference on Winter Simulation*, pp. 1211–1216. IEEE, Orlando, FL.
- Rothman, A.J., Bickel, P.J., Levina, E. and Zhu, J. (2008) Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, **2**, 494–515.
- Sackett, P.J., Al-Gaylani, M., Tiwari, A. and Williams, D. (2006) A review of data visualization: Opportunities in manufacturing sequence management. *International Journal of Computer Integrated Manufacturing*, **19**(7), 689–704.
- Salvendy, G. (2012) *Handbook of Human Factors and Ergonomics*. John Wiley & Sons, Hoboken, NJ.
- Song, P. X.-K., Li, M. and Yuan, Y. (2009) Joint regression analysis of correlated data using Gaussian copulas. *Biometrics*, **65**(1), 60–68.
- Tibshirani, R. (1996) Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**(1), 267–288.
- Wang, S., Parsons, M., Stone-McLean, J., Rogers, P., Boyd, S., Hoover, K., Meruvia-Pastor, O., Gong, M. and Smith, A. (2017) Augmented reality as a telemedicine platform for remote procedural training. *Sensors*, **17**(10), 2294.
- 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**(2), 66–77.
- Yuan, M. (2010) High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, **11**, 2261–2286.
- Yuan, M. and Lin, Y. (2007) Model selection and estimation in the Gaussian graphical model. *Biometrika*, **94**(1), 19–35.
- Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, **57**(298), 348–368.
- Zou, H. and Li, R. (2008) One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics*, **36**(4), 1509–1533.