

Article

Not peer-reviewed version

Bayesian Feature Extraction for Two-Part Latent Variable Model with Polytomous Manifestations

Qi Zhang , [Yi-Hui Zhang](#) , [And Yemao Xia](#) *

Posted Date: 1 February 2024

doi: [10.20944/preprints202401.2197.v2](https://doi.org/10.20944/preprints202401.2197.v2)

Keywords: Two-part latent variable model; spike and slab prior; Bayesian lasso; MCMC sampling; CHFS



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

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

Article

Bayesian Feature Extraction for Two-Part Latent Variable Model with Polytomous Manifestations

Qi Zhang¹, Yi-Hui Zhang² and Yemao Xia^{2,*}

¹ School of Science, Nanjing Forestry University, Nanjing, China; njfuzhangqi@njfu.edu.cn

² School of Computer Science, China University of Geosciences, Wuhan, China; yihuizhang@cug.edu.cn

* Correspondence: ymxia@njfu.edu.cn; Tel.: +86-25-8542-7703

Abstract: Semi-continuous data are very common in the social science and economics. In this paper, a Bayesian variable selection procedure is developed to assess the influence of exogenous factors including observed and unobserved on the semi-continuous data. Our formulation is based on the two-part latent variable model with polytomous response. We consider two schemes for the penalties of regression coefficients and factor loadings: the Bayesian spike and slab bimodal prior and the Bayesian lasso prior. Within the Bayesian framework, we implement Markov Chains Monte Carlo sampling method to conduct posterior inference. To facilitate posterior sampling, we recast the logistic model in part one as the norm-like mixture model. Gibbs sampler is designed to draw observations from the posterior. Our empirical results show that with suitable hyper-parameters, the spike and slab bimodal method slightly outperforms the Bayesian lasso in the current analysis. Finally, a real example related to the China household finance survey is analyzed to illustrate the application of the methodology.

Keywords: two-part latent variable model; spike and slab prior; Bayesian lasso, MCMC sampling; CHFS

MSC: 62H12; 62F15

1. Introduction

Semi-continuous data, characterized by excessive zeros, are very common in the fields of social science and economics. A typical example is given by [1] in the analysis of medical expenditures, in which the zeros correspond to a subpopulation of patients who do not use health services, while the positive values describe the actual levels of expenditures in use among users. In understanding such type of data structure, two-part model [2] is a widely appreciated statistical method. The basic assumption on two-part model is that the overall model is consisted of two processes: one binary process (Part one) and one continuous positive-valued process (Part two). The binary process, usually formulated within the logistic or probit regression model, is used to indicate whether the items are responded or not, while the continuous process, conditioning on the binary process, is used to describe what the actual levels of responses are (see, e.g., [3]). By combining two processes into one, two-part model provides a unified and flexible way in describing various relevance underlying semi-continuous data. Now, two-part model has been widely used in the health service [4–6], the medical expenditures [1,7–9], the household finance [10], the substance use study [11,12] and the genome analysis [13].

The traditional two-part model usually formulates the exogenous explanatory factors as fixed and observed. However, in the real applications especially in the social survey, many *unobserved/latent* and random factors also have important impacts on the outcome variable. This fact is revealed by [14] in the study of children's aggressive behavior. [14] noted that two factors, the propensity to engage in aggressive behavior and the propensity to have high aggressive activity levels, had significant influence on the children's aggressive behavior. They incorporated such two latent factors into analysis and established a two-component-two-part mixture model to identify the heterogeneity of population;

[15] noticed that in China, the finance literacy of a family had a nonignorable influence on the desire to holding finance debts, and also affected the amount of finance debt being held. They suggested conducting a joint analysis of latent factor and observed covariates in two-part regression model. The latent factor is further manifested by multiple binary measurements via factor analysis model. [16] incorporated two-part regression model into the general latent variable model framework and analyzed the internal relationships between multiple factors longitudinally. These methods have brought significant attention to the two-part model in behavioral science, economics, psychology, and medicine in recent years, see for example, [13,17,18] and references therein for further developments of two-part model.

In the analysis of semi-continuous data, an important issue is to determine which explanatory factor is helpful in improving model fit. This issue is especially true when the number of exogenous factors is large since the commonly used forward and backward regression procedure is extremely time-consuming. Now, lasso and its extensions [19–26] have been the most commonly used methods for the feature extraction. A typical feature of these methods is to put some suitable penalties on the coefficients and shrink many coefficients to zero, thus performing variable selection. Recently, these penalization/regularization approaches have been applied widely for prediction and prognosis (see for example, [27,28]). Though more appealing, the lasso-type regularization also suffers some limitations. For example, most contributions are developed within the frequency framework and the performance heavily depends on the large sample theory. It also readily leads to computational difficulty in the analysis of mixed data. An alternative for the variable selection is conducted within the Bayesian framework. Statisticians have introduced hierarchical models with mixture spike-and-slab priors that can adaptively determine the amount of shrinkage [29,30]. The spike-and-slab prior is the fundamental basis for most Bayesian variable selection approaches, and has proved remarkably successful [29–35]. Recently, Bayesian spike-and-slab priors have been applied to predictive modeling and variable selection in large scale genomic studies, see [36] for a simple review. Nevertheless, model selection has never been considered in the two-part regression model with latent variables. In this study, we introduce spike and slab model and Bayesian lasso into two-part latent variable model, which is the first attempt for this model.

Our formulation is more along with the lines of spike and slab bimodal prior in [33] and Bayesian Lasso in [37]. We formulate the problem by specifying a normal distribution with mean zero to the regression coefficient or factor loading of interest. The probability of a related variable being excluded or included is governed by the variance. To model the shrinkage of coefficients properly, we consider two schemes for the variance parameter: one is the two-point mixture model with one component located at the point close to zero and the other component situated at the point far away zero. The proportion is governed by a beta-distribution with suitable hyperparameters. Another scheme is along with the Bayesian lasso in which the variance is specified via a gamma distribution scaled by the penalty parameters. Two schemes are unified into a hierarchical mixture model. Within the Bayes paradigm, we developed a fully Bayesian selection procedure for the two-part latent variable model. We resort to the Markov Chains Monte Carlo sampling method. Gibbs sampler is used to draw observations from the posterior. We obtained all full conditionals. Posterior analysis is carried out based on the simulated observations. We investigate the performance of the proposed methods via simulation study and a real example. Our empirical results show that two schemes results in similar results in the variable selection but SS with suitable hyperparameters slightly outperforms over BaLsso in the correct rate.

The remainder of this paper is organized as follows. Section 2 introduces the proposed model for the semi-continuous data with latent variables. Section 3 develops the MCMC algorithm for the proposed model. Bayesian inference procedures including parameters estimation and model assessment are also presented in this section. In Section 4, we present the results of simulation study to assess the performance of our proposed methodology and illustrate the practical value of our proposed

model with household finance debt data. Section 5 concludes the paper with a discussion. Some technical details are given in the Appendix.

2. Model description

In section 2.1, a basic formulation for analyzing semi-continuous data with latent variables is presented. Section 2.2 presents a Bayesian procedure for the feature extraction.

2.1. Two-part latent variable model

Suppose that for $i = 1, 2, \dots, n$, s_i is a semicontinuous outcome variable which takes value in $[0, \infty)$; \mathbf{x}_i is a generic vector consisted of r fixed covariates representing the collection of *observed* explanatory factors of interest. We assume that each x_{ij} in \mathbf{x}_i is standardized in the sense $\sum_{i=1}^n x_{ij} = 0$ and $\sum_{i=1}^n x_{ij}^2 = 1$ for $j = 1, \dots, r$. Moreover, we include m *latent/unobserved* variables $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{im})^T$ into analysis to account for unobserved heterogeneity of responses. Conceptually, these latent variables can be the covariates that are not directly observed or the synthesization of some highly correlated explanatory items suffering from the noisy. Inclusion of latent variables can improve model fits and strengthen the power of model interpretation, see [38] for more discussions on the latent variables in a general setting. To deal with the spike of s_i at zero, we follow the common routine in literature (see for example, [9,11]) and identify s_i with two surrogate variables: $u_i = I\{s_i > 0\}$ and $z_i = \log(s_i^+)$, where $I(A)$ denotes the indicator function of set A and a^+ represents the positive part of a . That is, we separate the whole dataset into two parts: one part is the binary dataset which corresponds to the response-to-nonresponse indicators of subject and the other part is the logarithm of positive values. Our interest focuses on the exploration of effects of exogenous factors on two parts.

We assume that u_i and z_i satisfy the following sampling models:

$$p(u_i | \mathbf{x}_i, \boldsymbol{\omega}_i) = \frac{\exp(u_i \eta_i^u)}{1 + \exp(\eta_i^u)}, \quad (1)$$

$$\begin{aligned} \eta_i^u &= \alpha + \boldsymbol{\beta}_x^T \mathbf{x}_i + \boldsymbol{\beta}_\omega^T \boldsymbol{\omega}_i, \\ p(z_i | u_i = 1, \boldsymbol{\omega}_i) &= N(\eta_i^z, \sigma^2), \\ \eta_i^z &= \gamma + \boldsymbol{\psi}_x^T \mathbf{x}_i + \boldsymbol{\psi}_\omega^T \boldsymbol{\omega}_i, \end{aligned} \quad (2)$$

in which α and γ are the intercept parameters, $\boldsymbol{\beta}_x$ and $\boldsymbol{\psi}_x$ are the vectors of regression coefficients, and $\boldsymbol{\beta}_\omega$ and $\boldsymbol{\psi}_\omega$ are the vectors of factor loadings; σ^2 is the scale and ' T ' is the transpose operator of vector or matrix; For compactness, we write $\boldsymbol{\beta} = (\boldsymbol{\beta}_x^T, \boldsymbol{\beta}_\omega^T)^T$ and $\boldsymbol{\psi} = (\boldsymbol{\psi}_x^T, \boldsymbol{\psi}_\omega^T)^T$ and treat $\mathbf{w}_i = (\mathbf{x}_i^T, \boldsymbol{\omega}_i^T)^T$ as the complete explanatory variables.

The involvement of latent variables apparently complicates the model. It readily results in model identification problem [39,40]. This is especially true when the dimension of $\boldsymbol{\omega}_i$ is high. In this case, any auxiliary information is required to manifest $\boldsymbol{\omega}_i$ further. Among various easy-constructs, we consider latent variable (LV, [39,40]) approach. A basic assumption on LV approach is that there exist, say p manifestations $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})^T$, of which each y_{ij} may be continuous, counted or categorical, and assuming that they satisfy the following link equation

$$F(\mathbf{y}_i, \boldsymbol{\omega}_i, \boldsymbol{\epsilon}_i, \boldsymbol{\varphi}) = \mathbf{0}, \quad (3)$$

where F is a known and fixed link function, $\boldsymbol{\epsilon}_i$ is the vector of errors used to identify the idiosyncratic part of \mathbf{y}_i that can not be explained by $\boldsymbol{\omega}_i$, and $\boldsymbol{\varphi}$ is the vector of unknown parameters used to quantify the uncertainty of model. The information about $\boldsymbol{\omega}_i$ is manifested by \mathbf{y}_i via F . In this paper, in view of the real applications, we consider p ordered categorical variable $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})^T$, of which y_{ij} takes value in $\{0, 1, \dots, c_j\}$ ($c_j > 1$) and satisfies the following link model:

$$y_{ij} = \ell_j \quad \text{if} \quad \delta_{j\ell_j} < y_{ij}^* \leq \delta_{j,\ell_j+1}, \quad (4)$$

where $\delta_{j0} < \delta_{j1} < \dots < \delta_{jc_j} < \delta_{j,c_j+1}$ are the threshold parameters with $\delta_{j0} = -\infty$ and $\delta_{j,c_j+1} = +\infty$, and $\mathbf{y}_i^* = (y_{i1}^*, \dots, y_{ip}^*)^T$ is the vector of latent responses satisfying the factor analytic model:

$$\mathbf{y}_i^* = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \quad (5)$$

$$\boldsymbol{\omega}_i \stackrel{iid.}{\sim} N_m[\mathbf{0}, \boldsymbol{\Phi}], \quad \boldsymbol{\epsilon}_i \sim N_p[\mathbf{0}, \mathbf{I}_p], \quad \text{and } \boldsymbol{\omega}_i \perp \boldsymbol{\epsilon}_i, \quad (6)$$

where $\boldsymbol{\mu}$ is a p -dimensional intercept vector, $\boldsymbol{\Lambda}$ is the $p \times m$ -dimensional factor loading matrix, and \mathbf{I}_p is the identity matrix of order p . We assume that conditional upon $\boldsymbol{\omega}_i$, s_i and \mathbf{y}_i are independent.

We refer to the model specified by (1), (2) and (4) associated with (5) as the two-part latent variable model with polytomous responses. It provides a unified framework to explore the dependence of binary, continuous and categorical data simultaneously. The dependence between them results from the share of common factors or latent variables. If $\boldsymbol{\omega}_i$ is degenerated at zeros or the factor loadings are taken as zeros, the dependence among them disappears, and the overall model reduces to the traditional two-part model and ordinal regression model.

To facilitate the efficient calculation, motivated by the key identity in [41] (see equation (2) in their seminar paper), we express the logistics model (1) as the mixture model of form

$$\begin{aligned} & \frac{\exp(u_i(\alpha + \boldsymbol{\beta}^T \mathbf{w}_i))}{1 + \exp(\alpha + \boldsymbol{\beta}^T \mathbf{w}_i)} \\ &= 2^{-1} \exp(\kappa_i(\alpha + \boldsymbol{\beta}^T \mathbf{w}_i)) \int_0^\infty \exp\left\{-\frac{u_i^*}{2}(\alpha + \boldsymbol{\beta}^T \mathbf{w}_i)^2\right\} p_{PG}(u_i^*) du_i^*, \end{aligned} \quad (7)$$

where $\kappa_i = u_i - 0.5$, and $p_{PG}(u)$ is the standard Pólya-Gamma probability density function. Assuming that we introduce auxiliary variables u_i^* and augment them with u_i , then equation (1) can be considered as the marginal density of the joint distribution

$$p(u_i, u_i^* | \mathbf{x}_i, \boldsymbol{\omega}_i) = 2^{-1} \exp\left\{\kappa_i \eta_i^u - \frac{u_i^*}{2} \eta_i^{u2}\right\} p_{PG}(u_i^*). \quad (8)$$

Note that the exponential part in the brackets is the kernel of normal density function with respect to η_i^u . Hence, it admits conjugate full-conditional distributions for all regression coefficients, factor loadings and factor variables, leading to a straightforward Bayesian computation.

Let $\mathbf{U} = \{u_i\}_{i=1}^n$, $\mathbf{Z} = \{z_i\}_{i=1}^n$, and $\mathbf{Y} = \{y_i\}_{i=1}^n$ be the sets of observed variables; We write $\boldsymbol{\Omega} = \{\boldsymbol{\omega}_i\}_{i=1}^n$ for the collect of factor variables, and write $\mathbf{U}^* = \{u_i^*\}_{i=1}^n$, $\mathbf{V}^* = \{v_i^*\}_{i=1}^n$, $\mathbf{Y}^* = \{y_i^*\}_{i=1}^n$ for the sets of latent response variables. The complete-data likelihood is given by

$$\begin{aligned} & p(\mathbf{U}, \mathbf{U}^*, \mathbf{Z}, \mathbf{V}^*, \mathbf{Y}, \mathbf{Y}^*, \boldsymbol{\Omega} | \boldsymbol{\theta}) \\ &= p(\mathbf{U}, \mathbf{U}^* | \boldsymbol{\Omega}, \alpha, \boldsymbol{\beta}) p(\mathbf{Z}, \mathbf{V}^* | \mathbf{U}, \boldsymbol{\Omega}, \gamma, \boldsymbol{\psi}, \sigma^2) p(\mathbf{Y} | \mathbf{Y}^*, \delta) p(\mathbf{Y}^* | \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\boldsymbol{\Omega} | \boldsymbol{\Phi}) \\ &= \prod_{i=1}^n \exp\left\{\kappa_i \eta_i^u - \frac{1}{2} u_i^* (\eta_i^u)^2\right\} p_{PG}(u_i^* | 1, 0) \\ & \quad \times \prod_{i \in I} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2} (z_i - \eta_i^z)^2\right\} \\ & \quad \times \prod_{i=1}^n \prod_{j=1}^p \sum_{\ell=0}^{c_j} I\{\delta_{j\ell} < y_{ij}^* \leq \delta_{j,\ell+1}, y_{ij} = \ell\} \\ & \quad \times \prod_{i=1}^n \prod_{j=1}^p \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} (y_{ij}^* - \mu_j - \boldsymbol{\Lambda}_j^T \boldsymbol{\omega}_i)^2\right\} \\ & \quad \times \prod_{i=1}^n \frac{1}{(\sqrt{2\pi})^m |\boldsymbol{\Phi}|^{1/2}} \exp\left\{-\frac{1}{2} \text{tr}[\boldsymbol{\Phi}^{-1} \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T]\right\}. \end{aligned}$$

where $I = \{i : u_i = 1\}$ is the set of indices, $\delta = \{\delta_{j\ell}\}$ is the set of threshold parameters, and $\theta = \{\alpha, \beta, \gamma, \psi, \sigma^2, \mu, \Lambda, \Phi, \delta\}$ is the vector of unknown parameters. For the moment, we assume θ_j in θ all are free.

2.2. Bayesian feature selection

Generally speaking, regression variables \mathbf{x}_i and factor variables ω_i may not have impacts on the u_i and z_i simultaneously, and some redundant variables may exist. The presence of redundant variables not only decreases the model fit but also weakens the power of model interpretation. Therefore, it is necessary to determine which regression coefficient or factor loading is significantly away from zero. In the context of frequency statistics, this issue is generally tackled out via stepwise regression, in which each variable is decided to be exclude or included according to the model fit. However, the situation becomes complex when the number of independent variables is large. In this paper, we pursue a Bayesian variable selection procedure. To this end, we follow [37] and assume

$$\beta \sim N_q(\mathbf{0}, \text{diag}\{\gamma_{\beta k}^2\}), \quad \psi \sim N_q(\mathbf{0}, \sigma^2 \text{diag}\{\gamma_{\psi k}^2\}), \quad (9)$$

in which we use $\text{diag}\{a_k\}$ to represent a diagonal matrix with the k^{th} diagonal element a_k and let $q = r + m$. That is, we assume that each β_k in β (ψ_k is similar) is centered at zero (or equivalently each w_{ik} is excluded from \mathbf{w}_i) but with the probability governed by the variance $\gamma_{\beta k}^2$. If $\gamma_{\beta k}^2$ is close to zero, then the probability of β_k taking zero increases, and w_{ik} tends to be excluded; conversely, if $\gamma_{\beta k}^2$ is large, then the probability of β_k being zero is small and w_{ik} tends to be maintained. As a result, the value of $\gamma_{\beta k}^2$ plays a key role in determining whether w_k is relevant to be selected in Part one. With this in mind, a reasonable assumption on $\gamma_{\beta k}^2$ and $\gamma_{\psi k}^2$ is that:

$$\gamma_{\beta k}^2 \stackrel{\text{ind.}}{\sim} (1 - w_\beta) \delta_{\nu_{\beta 0} \eta_{\beta k}^2}(\cdot) + w_\beta \delta_{\eta_{\beta k}^2}(\cdot), \quad (10)$$

$$\gamma_{\psi k}^2 \stackrel{\text{ind.}}{\sim} (1 - w_\psi) \delta_{\nu_{\psi 0} \eta_{\psi k}^2}(\cdot) + w_\psi \delta_{\eta_{\psi k}^2}(\cdot), \quad (11)$$

where $\delta_a(\cdot)$ is the Dirac measure concentrated at point a , w_β is the random weight used to measure the similarity between $\gamma_{\beta k}^2$ and $\eta_{\beta k}^2$, and $\eta_{\beta k}^2$ is the hyperparameter used to represent how far β_k is away from zero or slab; $\nu_{\beta 0}$ is a previously specified small positive value used to identify the ‘spike’ of β_k at zero. In other words, every $\gamma_{\beta k}^2$ is assumed to be equal to $\eta_{\beta k}^2$ with probability w_β and equal to $\nu_{\beta 0} \eta_{\beta k}^2$ with probability $1 - w_\beta$. This is also true for w_ψ , $\eta_{\psi k}$ and $\nu_{\psi 0}$. To model w_β and w_ψ properly, we assign the following beta distributions to them

$$p(w_\beta | a_\beta, b_\beta) = \text{Beta}(a_\beta, b_\beta), \quad p(w_\psi | a_\psi, b_\psi) = \text{Beta}(a_\psi, b_\psi), \quad (12)$$

where a_β , a_ψ , b_β and b_ψ are the hyperparameters used to control the shape of beta density, that is, to determine the magnitude of weights in $(0, 1)$. For example, if $a_{\beta 1}$ in equation (12) is small and $b_{\beta 1}$ is large, then equation (12) encourages w_β to take small value with high probability. In contrast, it follows from $1 - \text{Beta}(a_\beta, b_\beta) = \text{Beta}(b_\beta, a_\beta)$ that large $a_{\beta 1}$ and small $b_{\beta 1}$ encourage w_β to take large value in $(0, 1)$. In the case that $a_{\beta 1} = b_{\beta 1} = 1.0$, equation (12) reduces to the uniform distributions on $(0, 1)$. In this case, every value in $(0, 1)$ is possible for w_β with the same probability. In the real applications, if no information can be available, one can assign the values to them to ensure the beta distribution to be inflated enough.

Finally, to measure the magnitudes of ‘slap’ in the distributions of β_k and ψ_k , we specify gamma distributions for $\eta_{\beta k}^{-2}$ and $\eta_{\psi k}^{-2}$, or equivalently,

$$\eta_{\beta k}^2 | a_{\beta 1}, a_{\beta 2} \stackrel{\text{iid.}}{\sim} IG(\tau_{\beta 0}, \zeta_{\beta 0}), \quad \eta_{\psi k}^2 | a_{\psi 1}, a_{\psi 2} \stackrel{\text{iid.}}{\sim} IG(\tau_{\psi 0}, \zeta_{\psi 0}), \quad (13)$$

where 'IG(a, b)' denotes the inverse-gamma distribution with mean $b/(a-1)$ for $a > 1$ and variance $b^2/((a-1)^2(a-2))$ for $a > 2$; $\tau_{\beta 0}$, $\zeta_{\beta 0}$, $\tau_{\psi 0}$ and $\zeta_{\psi 0}$ are the hyperparameters which are treated to fixed and known. Similarly, one can assign values to them to ensure (13) to be dispersed enough. For example, we can follow the routine in [33] in the ordinary regression analysis, and set $\tau_{\beta 0} = \tau_{\psi 0} = 1.0$ and $\zeta_{\beta 0} = \zeta_{\psi 0} = 0.05$ to obtain dispersed priors.

Note that equations (10) and (11) can be formulated as hierarchy as follows: for $k = 1, \dots, q$,

$$\gamma_{\beta k}^2 = f_{\beta k} \eta_{\beta k}^2, \quad \gamma_{\psi k}^2 = f_{\psi k} \eta_{\psi k}^2, \quad (14)$$

$$f_{\beta k} | \nu_{\beta 0}, \omega_{\beta} \stackrel{iid.}{\sim} (1 - w_{\beta}) \delta_{\nu_{\beta 0}}(\cdot) + w_{\beta} \delta_1(\cdot), \quad (14)$$

$$f_{\psi k} | \nu_{\psi 0}, w_{\psi} \stackrel{iid.}{\sim} (1 - w_{\psi}) \delta_{\nu_{\psi 0}}(\cdot) + w_{\psi} \delta_1(\cdot), \quad (15)$$

where $f_{\beta k}$ and $f_{\psi k}$ are the latent binary variables respectively. Such a formulation aims to separate $\eta_{\beta k}^2$ and $\eta_{\psi k}^2$ from the distributions (10) and (11) to facilitate posterior sampling.

It is instructive to compare the proposed method to the Bayesian lasso [37], in which the variance parameters $\gamma_{\beta k}^2$ and $\gamma_{\psi k}^2$ in equation (9) are specified via exponential distributions as follows:

$$p(\gamma_{\beta}^2 | \lambda_{\beta}^2) = \prod_{k=1}^q \frac{\lambda_{\beta k}^2}{2} \exp(-\lambda_{\beta k}^2 \gamma_{\beta k}^2 / 2), \quad (16)$$

$$p(\gamma_{\psi}^2 | \lambda_{\psi}^2) = \prod_{k=1}^q \frac{\lambda_{\psi k}^2}{2} \exp(-\lambda_{\psi k}^2 \gamma_{\psi k}^2 / 2), \quad (17)$$

where $\lambda_{\beta}^2 = (\lambda_{\beta 1}^2, \dots, \lambda_{\beta q}^2)^T$ and $\lambda_{\psi}^2 = (\lambda_{\psi 1}^2, \dots, \lambda_{\psi q}^2)^T$ are the shrinkage/penalty parameters used to control the amount of shrinkage of β_k and ψ_k toward zero.

Modeling $\gamma_{\beta k}^2$ and $\gamma_{\psi k}^2$ like equations (16) and (17) lead to marginal distributions of β_k and ψ_k as the laplace distributions with location zero and scale λ_k . The penalty parameters $\lambda_{\beta k}^2$ and $\lambda_{\psi k}^2$ are rather crucial in determining the amount of shrinkage of parameters. Figure 1 presents the plots of densities of Laplace distribution $LA(\lambda)(\lambda > 0)$ across various choices of λ . It can be seen that the larger the value of γ , the more kurtosis the density, indicating more penalties on the regression coefficient.

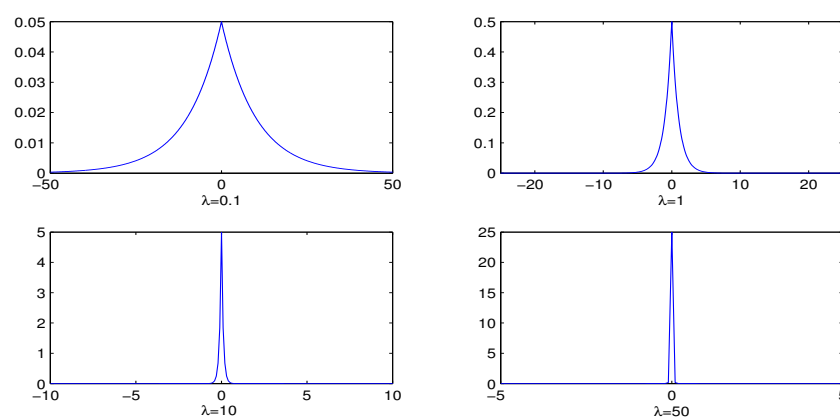


Figure 1. Plot of the densities of Laplace distribution across different choices of λ .

Due to their key role in equations (16) and (17), for λ_β^2 and λ_ψ^2 , we assign the following gamma priors to them, i.e.,

$$p(\lambda_\beta^2) = \prod_{k=1}^q p(\lambda_{\beta k}^2) = \prod_{k=1}^q Ga(a_{k0}, b_{k0}), \quad (18)$$

$$p(\lambda_\psi^2) = \prod_{k=1}^q p(\lambda_{\psi k}^2) = \prod_{k=1}^q Ga(c_{k0}, d_{k0}), \quad (19)$$

where ' $Ga(\nu, \lambda)$ ' denotes the gamma distribution with mean ν/λ . As the previous discussions, the values of a_{k0} , b_{k0} , c_{k0} and d_{k0} should be selected with care since they relate the shrinkages directly. Similar to that in (13), one can set $a_{k0} = c_{k0} = 1$, $b_{k0} = d_{k0} = 0.05$ to enhance the robustness of inference. This routine is followed in our empirical study.

Let $\mathbf{F}_\beta^* = \{f_{\beta k}\}$, $\mathbf{F}_\psi^* = \{f_{\psi k}\}$, $\gamma_\beta^2 = \{\gamma_{\beta k}^2\}$, $\gamma_\psi^2 = \{\gamma_{\psi k}^2\}$, $\eta_\beta^2 = \{\eta_{\beta k}^2\}$, $\eta_\psi^2 = \{\eta_{\psi k}^2\}$. We treat $\nu_{\beta 0}$ and $\nu_{\psi 0}$ as known hyperparameters. Note that γ_β^2 and γ_ψ^2 are totally determined by \mathbf{F}_β^* , \mathbf{F}_ψ^* and η_β^2 , η_ψ^2 . In the following, we abbreviate spike and slab bimodal prior to SS and Bayesian lasso to BaLsso.

3. Bayesian inference

3.1. Prior specification and MCMC sampling

In view of the model complexity, we consider Bayesian inference. Some priors are required to specify for unknown parameters to complete Bayesian model specification. Based on the model convention, it is naturally to assume that the parameters involved in the different models are independent.

Firstly, for μ , Λ and Φ , we consider the following conjugate priors:

$$p(\mu) = N_p(\mu_0, \Sigma_0), \quad (20)$$

$$p(\Lambda) = \prod_{k=1}^p p(\Lambda_k) = \prod_{k=1}^p N_m(\Lambda_{0k}, \mathbf{H}_{0k}), \quad (21)$$

$$p(\Phi) = IW(\rho_0, \mathbf{R}_0^{-1}), \quad (22)$$

where ' $IW(\rho, \mathbf{R})$ ' denotes the inverse Wishart distribution with degrees of freedom ρ and scale matrix \mathbf{R} [42]; Λ_k^T is the k^{th} row vector of Λ ; $\mu_0, \Sigma_0(p \times p) > 0$, $\Lambda_{0k}, \mathbf{H}_{0k}(m \times m) > 0$, $\rho_0 > 0$, and $\mathbf{R}_0(m \times m) > 0$ are the hyperparameters which are treated to be fixed and known.

Secondly, for α , γ , σ^2 in part one and two, we assume they are mutually independent and satisfy

$$p(\alpha) = N(\alpha_0, \sigma_{\alpha 0}^2), \quad p(\gamma) = N(\gamma_0, \sigma_{\gamma 0}^2), \quad p(\sigma^{-2}) \sim Ga(a_0, b_0), \quad (23)$$

where α_0 , $\sigma_{\alpha 0}^2$, γ_0 , $\sigma_{\gamma 0}^2$ and a_0 , b_0 are the fixed hyperparameters.

Lastly, for threshold parameter δ , without loss of generality, we assume that c_j , the number of categories of y_{ij} , is invariant across the subscript j and equals to c . Moreover, we assume that $p(\delta) = \prod_{j=1}^p p(\delta_j)$, where $\delta_j = (\delta_{jk})$ is the j^{th} row vector of δ . In the following, we suppress the subscript j in δ_{jk} for notational simplicity and write δ for δ_j .

Let $F_0(\cdot)$ be any strictly monotonically increasing and differentiable function on \mathbb{R} with $F_0(+\infty) = 1$ and $F_0(-\infty) = 0$. For example, one can take $F_0 = \Phi(\cdot/\tau_0)$ for some $\tau_0 > 0$ or student distribution with degrees of freedom ν_0 , where $\Phi(\cdot)$ is the standard normal distribution function. To specify a prior for δ , we follow [43] and let $p_j = F_0(\delta_j) - F_0(\delta_{j-1})$ for $j = 1, \dots, c$. It is easily to show that the

transformation is invertible with Jacobi determination unity. We first consider the following Dirichlet distribution for $\mathbf{p} = (p_1, \dots, p_c)^T$:

$$\pi(\mathbf{p}) = \frac{1}{\mathcal{B}(\eta_1, \dots, \eta_{c+1})} p_1^{\eta_1-1} \dots p_c^{\eta_c-1} (1 - \sum_{\ell=1}^c p_\ell)^{\eta_{c+1}-1}$$

where $\mathcal{B}(\eta_1, \dots, \eta_{c+1}) = \prod_{j=1}^{c+1} \Gamma(\eta_j) / \Gamma(\sum_{j=1}^{c+1} \eta_j)$ is the multivariate beta function evaluated at $\eta_1, \dots, \eta_{c+1}$, and $\eta_j > 0$. Then, by the formula of inverse transformation, the joint distribution of δ is given by

$$\pi(\delta) = \frac{1}{\mathcal{B}(\eta_1, \dots, \eta_{c+1})} p_1^{\eta_1-1} \dots p_c^{\eta_c-1} (1 - \sum_{\ell=1}^c p_\ell)^{\eta_{c+1}-1} \prod_{j=1}^c f_0(\delta_j) I\{\delta_1 < \dots < \delta_c\}, \quad (24)$$

where $f_0(x)$ is the derivative of $F_0(x)$ with respect to x . We call (24) the transformed Dirichlet prior and use it as the prior of δ . An advantage of working with (24) is that conditional upon δ_{j-1} and δ_{j+1} , the transformed distribution of δ_j has the beta distribution given by

$$\frac{F_0(\delta_j) - F_0(\delta_{j-1})}{F_0(\delta_{j+1}) - F_0(\delta_{j-1})} | (\delta_{j-1}, \delta_{j+1}) \sim \text{Beta}(\eta_j, \eta_{j+1}), (j = 1, \dots, c). \quad (25)$$

3.2. MCMC sampling

With the prior given above, the inference about θ is based on the posterior $p^o(\theta | \mathbf{U}, \mathbf{Z}, \mathbf{Y})$, which has no closed form. Motivated by the key idea in [44], we treat latent quantities as the missing data and argument them to the observed data to form the complete data. The statistical inference is carried out based on the complete-data likelihood. For this end, apart from Ω , \mathbf{U}^* and \mathbf{Y}^* mentioned before, we further let \mathbf{Q}^* be the collection of latent quantities involved in the specifications of β and ψ , i.e., $\mathbf{Q}^* = \{\mathbf{F}_\beta^*, \mathbf{F}_\psi^*, \eta_\beta^2, \eta_\psi^2, w_\beta, w_\psi\}$ under SS and $\{\lambda_\beta^2, \lambda_\psi^2\}$ under BaLsso. Rather than working with the posterior p^o directly, we consider the following joint distribution

$$p^{joint}(\Omega, \mathbf{U}^*, \mathbf{Y}^*, \mathbf{Q}^*, \theta | \mathbf{U}, \mathbf{Z}, \mathbf{Y}), \quad (26)$$

where p^o can be considered as the marginal of p^{joint} . We use Markov chain Monte Carlo(MCMC, [45,46]) sampling method to simulate observations from this target distribution. In particular, Gibbs sampler is implemented to draw observations iteratively from the full conditional distributions as follows:

- draw Ω from $p(\Omega | \mathbf{U}^*, \mathbf{Q}^*, \mathbf{Y}^*, \theta, \mathbf{U}, \mathbf{Z}, \mathbf{V})$,
- draw \mathbf{U}^* from $p(\mathbf{U}^* | \Omega, \mathbf{Y}^*, \mathbf{Q}^*, \theta, \mathbf{U}, \mathbf{Z}, \mathbf{V})$,
- draw \mathbf{Y}^* from $p(\mathbf{Y}^* | \Omega, \mathbf{U}^*, \mathbf{Q}^*, \theta, \mathbf{U}, \mathbf{Z}, \mathbf{V})$,
- draw \mathbf{Q}^* from $p(\mathbf{Q}^* | \Omega, \mathbf{U}^*, \mathbf{Y}^*, \theta, \mathbf{U}, \mathbf{Z}, \mathbf{V})$, and
- draw θ from $p(\theta | \Omega, \mathbf{U}^*, \mathbf{Y}^*, \mathbf{Q}^*, \mathbf{U}, \mathbf{Z}, \mathbf{V})$.

Upon convergence, the posterior is approximated by the empirical distribution of the simulated observations. The convergence of algorithm can be monitored by plotting the traces of estimates under different starting values or observing the values of EPSR [47] of unknown parameters. The technical details on implementing MCMC sampling are given in Appendix.

Simulated observations obtained from the blocked Gibbs sampler can be used for statistical inference via a straightforward analysis procedure. For example, the joint Bayesian estimates of unknown parameters can be obtained via sample averaging as follows:

$$\hat{\theta} = M^{-1} \sum_{m=1}^M \theta^{(m)},$$

where $\{\theta^{(m)} : m = 1, \dots, M\}$ are the simulated observations from the posterior. The consistent estimates of covariance matrices of estimates can be obtained via sample covariance matrices.

The main purpose of introducing SS and BaLsso is to screen the variable in \mathbf{w}_i . Unlike that in the frequency statistics, Bayesian variable selection does not produce the estimates $\hat{\beta}$ and $\hat{\psi}$ exactly equal to zero, and hence it is necessary to determine which component can be treated as zero. This can be accomplished via posterior confidence intervals (PCI) of β_j and ψ_j , given by

$$P(|\beta_j| < c_{\alpha/2} | \mathbf{U}, \mathbf{Z}, \mathbf{Y}) = 1 - \alpha, \quad P(|\psi_j| < d_{\alpha/2} | \mathbf{U}, \mathbf{Z}, \mathbf{Y}) = 1 - \alpha \quad (27)$$

where α is any previously specified value in $(0, 1)$. The calculation of PCI can be achieved via Monte Carlo method. For example, let $\beta_j^{(k)} : k = 1, \dots, K$ be the K observations generated from the posterior distribution, then the PCI of β_j with confidence level $100(1 - \alpha)\%$ is given by $[\beta_{j,100(\alpha/2)}, \beta_{j,100(1-\alpha/2)}]$, where $\beta_{j,k}$ is the k^{th} order statistics.

Another choice for variable determination in SS is based on the posterior probability of $f_{\beta_j} = 1$ and $f_{\psi_j} = 1$, which can be approximated by

$$\hat{f}_{\beta_j} = \frac{1}{K} \sum_{k=1}^K I\{f_{\beta_j}^{(k)} = 1\}, \quad \hat{f}_{\psi_j} = \frac{1}{K} \sum_{k=1}^K I\{f_{\psi_j}^{(k)} = 1\}, \quad (28)$$

where $f_{\beta_j}^{(k)}$ and $f_{\psi_j}^{(k)}$ ($k = 1, \dots, K$) are the k observations drawn from the posterior distribution via Gibbs sampler. The variable w_j is selected in part one and two if $\hat{f}_{\beta_j} > 0.5$ and $\hat{f}_{\psi_j} > 0.5$.

4. Simulation study

In this section, a simulation study is conducted to assess the performance of the proposed method. The main objective is to assess the accuracy of estimates and the correct rate of variable selection. We consider one semi-continuous variable s_i , two factor variables ω_{i1} and ω_{i2} , and six categorical variables y_{ij} ($j = 1, \dots, 6$). We assume that s_i , ω_{ij} and y_{ij} satisfy equations (1), (2) and (4) associated with (5), respectively, in which the number of fixed covariates is set at five. We generate x_{i1} and x_{i2} from the standard normal distribution, x_{i3} and x_{i4} from the binomial distribution with probability of success 0.3, and x_5 from the uniform distribution on $(0, 1)$. All covariates were standardized to unify the scales. For ordered categorical variables, we take $c_j = c = 4$, that is, each y_{ij} belongs to $\{0, 1, 2, 3, 4\}$.

The true values of population parameters are set as follows: $\alpha = \gamma = 0.7$, $\beta = (0.7, 0.0, 0.7, 0.0, 0.7, 0.0, 0.8)^T$, $\gamma = 0.7$, $\psi = (0.7, 0.0, 0.7, 0.0, 0.7, 0.8, 0.0)^T$, $\sigma^2 = 1.0$, $\mu = 0.7 \times \mathbf{1}_6$, in which $\mathbf{1}_6$ is a 6×1 vector with elements being unity. The factor loading matrix Λ and covariance matrix Φ are taken as

$$\Lambda^T = \begin{bmatrix} 1.0 & 0.8 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.8 & 0.8 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}, \quad (29)$$

in which ones and zeros in Λ are treated as fixed to identify model; the thresholds are set as $\delta_k = (-1.5^*, 0.0, 1.2, 2.5^*)^T$ for $k = 1, \dots, 6$, where the elements with an asterisk are treated as fixed for model identification. Based on these setups, we generate data by first drawing latent factors from $N_2(\mathbf{0}, \Phi)$, and then drawing latent responses \mathbf{Y}^* from (5). The ordered categorical responses \mathbf{Y} , the indicator responses \mathbf{U} and the intensity responses \mathbf{Z} are sequentially generated from (5), (1) and (2). To investigate the effect of sample size on the estimates, we take $n = 400$ and 1000 , respectively, which represent small and large levels of sample size.

For Bayesian analysis, we consider the following inputs for hyperparameters: for the parameters involved in the measurement model, we take $\mu_0 = \mathbf{0}_6$, and $\Sigma_0 = 100.0 \times \mathbf{I}_6$; the elements in Λ_0 corresponding to the free parameters in Λ are set at zero, and $\mathbf{H}_{0k} = \mathbf{I}_2$ for $k = 1, \dots, 6$; $\rho_0 = 10.0$, and $\mathbf{R}_0^{-1} = 6.0 \times \mathbf{I}_3$; for the threshold parameters δ , we take $\eta_1 = \dots = \eta_5 = 1.0$, which denotes the

uniform distribution of \mathbf{p} on the simplex in \mathbb{R}^5 ; for intercept parameters α , γ and σ^2 in the two-part model, we set $\alpha_0 = \gamma_0 = 0$, $\sigma_{\alpha 0}^2 = \sigma_{\gamma 0}^2 = 100$, and $a_0 = b_0 = 2.0$; the hyperparameters involved in the formulation of β and ψ are set as before. Note that these values can ensure the priors inflated enough, hence it could be expected to enhance the robust of inference. In addition, we set $\nu_{\beta 0} = \nu_{\psi 0} = 0.001$ in equations (10) and (11) to guarantee β_k and ψ_k clumping at zero sufficiently.

The MCMC algorithm described in Section 3 is implemented to obtain the estimates of unknown parameters θ . Before formal implementation, a few test runs were conducted as pilots to monitor the convergence of the Gibbs sampler. We plot the values of EPSR of unknown parameters against the number of iterations under three different starting values. For SS, Figure 2 presents the plots of EPSR of unknown parameters under three different starting value with sample size $n = 400$.

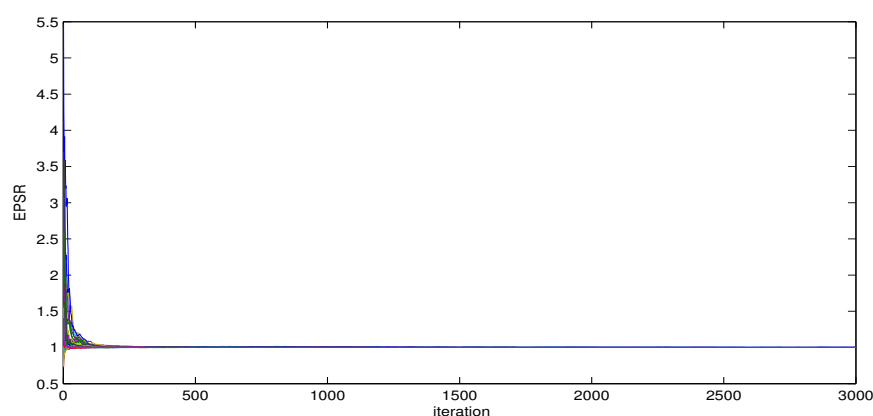


Figure 2. Plot of the values of EPSR of unknown parameters under three different starting values: simulation study and $n = 400$.

It can be found that the convergence of estimates is fast and all values of EPSR are less than 1.2 in about 300 iterations. To be conservative, we remove the first 2000 observations as burn-in phrase and further collect 3000 observations for calculating the bias (BIAS), the root mean squares (RMS) and the standard deviation (SD) of the estimate across 100 replications. The BIAS and RMS of the j -th component $\hat{\theta}_j$ in estimates are defined as follows:

$$\text{BIAS}(\hat{\theta}_j) = (\bar{\theta}_j - \theta_j^0), \bar{\theta}_j = \frac{1}{100} \sum_{\kappa=1}^{100} \hat{\theta}_j^{(\kappa)}, \text{RMS}(\hat{\theta}_j) = \sqrt{\frac{1}{100} \sum_{\kappa=1}^{100} (\hat{\theta}_j^{(\kappa)} - \theta_j^0)^2}, \quad (30)$$

where θ_j^0 is the j -th element of population parameters θ^0 . The summaries of estimates of main parameters under two scenarios are reported in Tables 1 and 2, where the sums of SD and RMS across the estimates are presented in the last rows.

Table 1. Summary of the estimates of unknown parameters under SS and BaLsso: simulation study and $n = 400$.

PAR	SS			BaLsso		
	BIAS	RMS	SD	BIAS	RMS	SD
$\alpha = 0.7$	-0.015	0.097	0.129	0.028	0.150	0.134
$\beta_1 = 0.7$	-0.056	0.143	0.142	-0.152	0.217	0.136
$\beta_2 = 0.0$	-0.001	0.021	0.061	-0.019	0.042	0.079
$\beta_3 = 0.7$	-0.144	0.216	0.145	-0.122	0.251	0.148
$\beta_4 = 0.0$	0.005	0.030	0.064	-0.008	0.040	0.078
$\beta_5 = 0.7$	-0.091	0.147	0.137	-0.045	0.135	0.137
$\beta_6 = 0.0$	0.017	0.028	0.075	0.026	0.055	0.096
$\beta_7 = 0.8$	-0.187	0.237	0.184	-0.126	0.209	0.184
$\gamma = 0.7$	0.010	0.079	0.084	0.008	0.063	0.085
$\psi_1 = 0.7$	-0.035	0.079	0.077	-0.011	0.065	0.074
$\psi_2 = 0.0$	0.005	0.032	0.051	-0.018	0.031	0.054
$\psi_3 = 0.7$	-0.007	0.061	0.070	-0.021	0.085	0.069
$\psi_4 = 0.0$	-0.007	0.029	0.049	-0.003	0.031	0.053
$\psi_5 = 0.7$	-0.070	0.093	0.077	-0.018	0.082	0.075
$\psi_6 = 0.8$	-0.040	0.086	0.089	-0.02	0.069	0.088
$\psi_7 = 0.0$	-0.011	0.033	0.062	0.014	0.036	0.069
$\sigma^2 = 1.0$	0.085	0.129	0.117	0.038	0.082	0.111
$\lambda_{21} = 0.8$	0.042	0.078	0.073	0.058	0.098	0.071
$\lambda_{31} = 0.8$	0.030	0.072	0.071	0.034	0.063	0.072
$\lambda_{52} = 0.8$	0.058	0.079	0.072	0.052	0.090	0.073
$\lambda_{62} = 0.8$	0.031	0.060	0.072	0.037	0.064	0.073
$\Phi_{12} = 0.3$	0.014	0.041	0.074	0.018	0.058	0.076
Total	-	1.870	1.975	-	2.016	2.035

Table 2. Summary of the estimates of unknown parameters under SS and BaLsso: simulation study and $n = 1000$.

PAR	SS			BaLsso		
	BIAS	RMS	SD	BIAS	RMS	SD
$\alpha = 0.7$	0.052	0.096	0.087	0.009	0.092	0.087
$\beta_1 = 0.7$	0.005	0.069	0.089	0.055	0.117	0.090
$\beta_2 = 0.0$	0.003	0.048	0.058	0.032	0.052	0.060
$\beta_3 = 0.7$	0.007	0.086	0.093	-0.045	0.076	0.091
$\beta_4 = 0.0$	0.004	0.015	0.049	-0.020	0.043	0.060
$\beta_5 = 0.7$	0.010	0.071	0.086	0.013	0.074	0.085
$\beta_6 = 0.0$	-0.003	0.029	0.059	0.032	0.064	0.077
$\beta_7 = 0.8$	0.002	0.102	0.120	-0.042	0.108	0.114
$\gamma = 0.7$	0.017	0.042	0.053	0.030	0.056	0.054
$\psi_1 = 0.7$	-0.023	0.038	0.046	-0.016	0.039	0.047
$\psi_2 = 0.0$	-0.007	0.019	0.033	-0.005	0.018	0.037
$\psi_3 = 0.7$	-0.028	0.060	0.042	-0.014	0.026	0.043
$\psi_4 = 0.0$	-0.007	0.023	0.033	0.000	0.018	0.036
$\psi_5 = 0.7$	-0.005	0.035	0.046	0.003	0.043	0.047
$\psi_6 = 0.8$	-0.031	0.058	0.053	-0.039	0.063	0.054
$\psi_7 = 0.0$	-0.001	0.031	0.045	-0.025	0.081	0.053
$\sigma^2 = 1.0$	0.018	0.049	0.068	0.041	0.053	0.071
$\lambda_{21} = 0.8$	0.021	0.041	0.045	0.033	0.038	0.045
$\lambda_{31} = 0.8$	0.016	0.049	0.045	0.028	0.038	0.045
$\lambda_{52} = 0.8$	0.032	0.049	0.045	0.054	0.057	0.045
$\lambda_{62} = 0.8$	0.043	0.059	0.046	0.043	0.054	0.046
Φ_{12}	0.016	0.043	0.049	0.005	0.037	0.048
Total	-	1.112	1.29	-	1.247	1.335

Examinations of Tables 1 and 2 present the following findings: (i) Both methods produce satisfactory results and the performance of SS are slightly superior to that of BaLsso. For $n = 400$, the total RMS and SD are 1.870 and 1.975 respectively under SS, and amount to 2.016 and 2.035 respectively under BaLsso; (ii) As expected, increasing the sample size improves the accuracy of the estimates both for SS and BaLsso.

Another simulation is conducted to assess the performance of the proposed method in the variable selection when the covariates and latent variables are correlated. In this setting, we generate covariates and latent factors jointly from the multivariate normal distribution with mean zeros and covariance matrix $\Sigma(7 \times 7)$ with $\Sigma_{jk} = \rho^{|j-k|}$, where Σ_{jk} is the $(j, k)^{th}$ entry of Σ . We consider three scenarios for ρ : (i) $\rho = 0.1$, (ii) $\rho = 0.5$ and (iii) $\rho = 0.8$, which represents respectively the weak, the mild and the strong dependence among them. The values of β and ψ are taken as $(1.0, 0.0, 1.0, 0.0, 1.0, 0.0, 1.0)$ and $(1.0, 0.0, 1.0, 0.0, 1.0, 1.0, 0.0)$ respectively, and the sample size is taken as $n = 1000$. The other model setups are set as the same as before. We implement MCMC sampling and collect 3000 observations after removing first 2000 observations for posterior inference. We follow [48] and treat a regression coefficient to be zero if the absolute value of its estimate is less than 0.1. Table 3 gives the summary of variable selection across 100 replications.

Table 3. Number of correctly selected variables in the two-part model on the simulated data sets.

PAR	SS			BaLsso		
	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.8$
$\beta_1 = 1.0$	100	100	100	100	100	100
$\beta_2 = 0.0$	98	96	85	88	86	76
$\beta_3 = 1.0$	100	100	100	100	100	100
$\beta_4 = 0.0$	96	95	86	93	93	85
$\beta_5 = 1.0$	100	100	100	100	100	100
$\beta_6 = 0.0$	96	94	93	97	92	87
$\beta_7 = 1.0$	100	100	100	100	100	100
$\psi_1 = 1.0$	99	100	100	100	100	100
$\psi_2 = 0.0$	100	99	95	100	98	93
$\psi_3 = 1.0$	100	100	100	100	100	100
$\psi_4 = 0.0$	100	100	97	98	100	91
$\psi_5 = 1.0$	100	100	100	100	100	100
$\psi_6 = 1.0$	100	100	100	100	100	100
$\psi_7 = 0.0$	100	98	97	97	96	96

Based on Table 3, it can be found that (i) for nonzero regression coefficients, two methods exhibit satisfactory performances, both with 100% correct rates across all situations; (ii) for zero regression coefficients, there exist difference between two methods, and SS are uniformly outperforms BaLsso. The underlying reason perhaps is that for SS, the variances of estimates are set to be small enough to ensure the coefficients close to zero while for BaLsso, the variance of estimates are controlled by the shrinkage parameters which may not be large enough to ensure this point; (iii) with the increase of the strength of dependence, the correct rates of two methods decreases.

5. China Household Finance Survey data

To demonstrate the usefulness of the proposed methodology, in this section a small portion of Chinese household finance debt data is analyzed. The dataset is collected from the China household financial survey (CHFS), a non-profit institute organized by the Southeast University of Finance and Economics. The survey covers a series of questions which touch on the information about various aspects of the household’s financial situation. In this study, we only focus on the measurement ‘gross debts per household (DEB)’, the amount of the secured debt and unsecured debt of a household under investigation. We extracted them from the survey of Zhejiang Province in 2013. Due to some uncertain

factors, some measurements in DEB are missing. The missing proportion is about 2.7%. We remove the subjects with missing entries and the ultimate sample size is 884. A preliminary data analysis shows that the measurements DEB contain excessive zeros and the proportion of zeros is about 72.58%. Naturally, we treat this variable as the outcome variable s_i , and identify it with u_i and z_i . Figure 3 presents the histogram of DEB as well as the logarithms of positive values. It can be seen clearly that dataset illustrates strong heterogeneity. The skewness and kurtosis of DEB are 1.1042 and 2.3361, respectively, which indicates that single parametric model for DEB may be unappreciated.

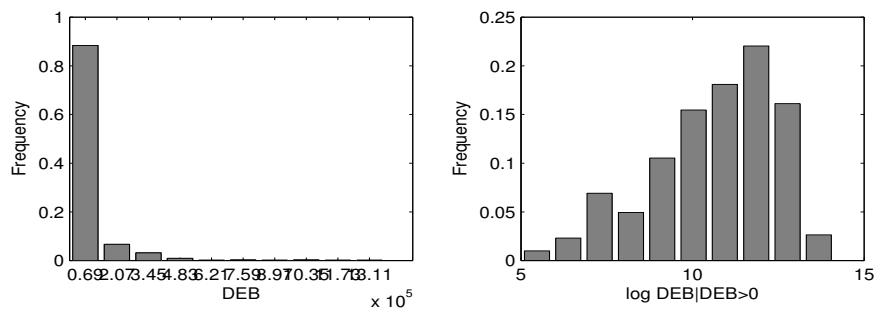


Figure 3. Histograms of DEB and the logarithms of their positive values: China household finance survey data. Left panel corresponding the DEB and right panel corresponding to $\log(\text{DEB}|\text{DEB} > 0)$.

We include the following measurements as the potential explanatory factors to interpret the variability in DEB: gender (x_1), age (x_2), marital status (x_3), health condition (x_4), educational experience (x_5), employment status of the household head (x_6), the number of family members (aged over 16, x_7), and the household annual income (x_8). Table 4 gives the descriptive summary of the measurements under consideration. To unify the scale, all covariates were standardized.

Table 4. Descriptive statistics of explanatory variables: CHFS data .

Variable.	Description.	Mean.	Max.	Min.	SD
Gender (x_1)	=1, male; =0, otherwise	0.756	1	0	0.430
Age (x_2)		51.81	91	19	14.931
Marital status (x_3)	=1, married; 0, otherwise	0.863	1	0	0.344
Health condition (x_4)	=1, good; 0, otherwise	0.833	1	0	0.373
Education degree (x_5)	=1, high school or above; =0, otherwise	0.352	1	0	0.478
Employment (x_6)	=1, yes; 0, otherwise	0.092	1	0	0.290
No. of adults (x_7)		3.002	3	0	1.301
Annual Income (CYN)(x_8)*		9.376 ⁴	8.060 ⁵	0	4.249 ⁴

* Note: Superscripts are used to indicate values raised to the power 10 (thus $a^b = a \times 10^b$). The measurement is taken as the middle value of the range in the questionnaire.

Besides the observed factors mentioned above, we also include family culture η , a latent factor into current analysis. It is well-known that China is an ancient civilization country with a long history, and Confucian culture has deeply rooted in the social development. Economic activity or social development can not be independent of cultural development. Hence, it is of practical interest to investigate how the family culture affect the behaviour of the household finance debt. Based on the design of the questionnaire, we select the following three measurements as manifestations for η : (i) boys's preference (BP, y_1). This is a three-category measurement coded by 0, 1 and 2, which corresponds to the attitude 'oppositive', 'does't matter' and 'strongly support'; (ii) attitude toward to the single of children (SC), coded by 0, 1 and 2, according to the leve of support; (iii) importance of a household head in a family. This measurement is originally coded in point 0 to 5 according to the support level. However, in view of that the frequencies in the last three groups are small, we group them into three

categories and recode them by 0 (does not matter), 1(important) and 2 (very important). In addition, due to that some manifestations are missing, we treat missing data as missing random and ignorable [49], and ignore the specific missing mechanic that results in missing data.

Let $\mathbf{U} = \{u_i\}$, $\mathbf{Z} = \{z_i\}$, and $\mathbf{Y} = \{\mathbf{Y}_{obs}, \mathbf{Y}_{mis}\}$, where \mathbf{Y}_{obs} is the collectionn of observed data and \mathbf{Y}_{mis} is the set of missing data. We formulate \mathbf{U} , \mathbf{Z} and \mathbf{Y} within equations (1), (2) and (5), and assume that $\eta_i, iid. \sim N(0,1)$. The inputs of hyperparameters in the priors are taken as follows: $\Lambda_{j0} = 0.0$, $H_{j0} = 1$ and $\eta_{j1} = \eta_{j2} = \eta_{j3} = 2.0$. The values of other hyperparameters are taken as the same as those in the simulation study. To implement MCMC sampling algorithm, we need to impute the missing data in \mathbf{Y} . This is just to do by drawing $y_{ij,mis}$ from the conditional distribution $p(y_{ij,mis}|\boldsymbol{\theta}, \mathbf{Y}_{obs}) = N(\mu_{j,mis} + \Lambda_{j,mis}\eta_i, 1)$, where $\mu_{j,mis}$ and $\Lambda_{j,mis}$ are the components of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ respectively which corresponds to the missing entry $y_{ij,mis}$ in \mathbf{y}_i . In addition, to identify the model and scale the factor, we set $\Lambda_1 = 1$. We also adopt the method in [50] in the context of latent variable model with polytomous data and fix δ_{j1} at $\Phi^{-1}(f_{j1}/n_j)$, where n_j is the size of $y_{obs,ij}$ equal to 1, and f_{j1} is the observed frequency of 0 in $y_{obs,ij}$. To assess the convergence of the algorithm, for SS, we plot the traces of estimates under three different initial values (see Figure 4). It can be seen that the algorithm converges at about 3000 iterations. To be conservative, we collect 6000 observations after deleting the initial 4000 observations for calculating the estimates and the standard deviations.

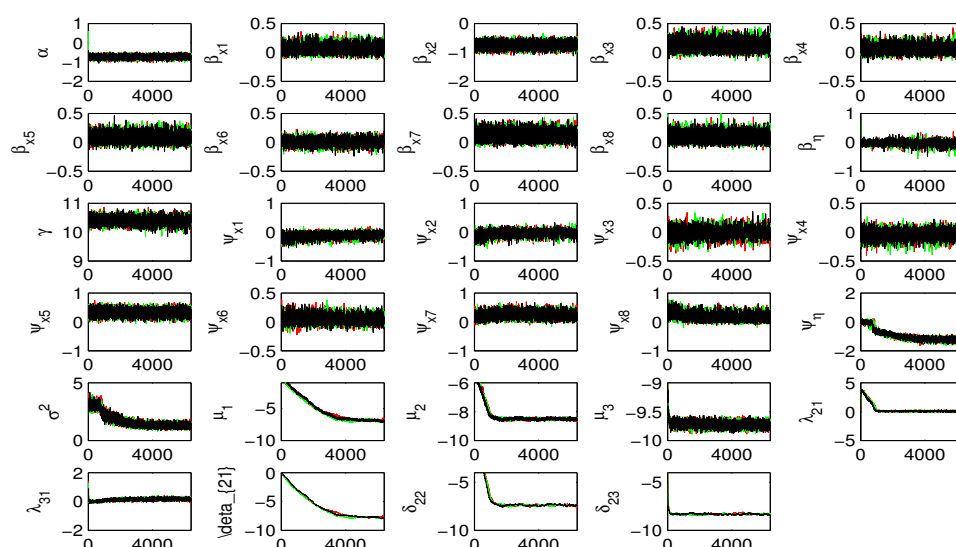


Figure 4. Trace plots of the estimates of unknown parameters against the number of iteration under SS prior: CHFS data.

Table 6 gives the summary of the estimates of unknown parameters in two parts and factor loadings. Examinations of Table 6 show that most estimates are very close but there exists differences in the estimates of β_4 , β_5 , β_7 , β_8 , ψ_2 , ψ_7 and ψ_8 . For example, the estimates of β_4 , β_5 and β_7 under SS are 0.428, 0.577, 0.747 with standard deviations 0.062, 0.070 and 0.072 respectively, while equal to 0.072, 0.082 and 0.092 with standard deviations 0.07, 0.081 and 0.092 under BaLsso. These differences reflect the fact that two methods impose different penalties on the regression coefficients in the variable selection.

Table 5. Estimates and standard deviations estimates of unknown parameters under SS and BaLsso: CHFS data.

SS					BaLsso				
Par	Est.	SD	Est.	SD	Par	Est.	SD	Est.	SD
α	-0.835	0.078	-0.838	0.080	γ	9.782	0.152	9.670	0.125
β_1	0.050	0.063	0.076	0.070	ψ_1	-0.137	0.103	-0.107	0.088
β_2	-0.750	0.099	-0.757	0.102	ψ_2	-0.147	0.141	-0.015	0.081
β_3	0.107	0.085	0.147	0.088	ψ_3	-0.022	0.065	-0.006	0.075
β_4	0.428	0.062	0.072	0.070	ψ_4	-0.019	0.060	-0.029	0.069
β_5	0.577	0.070	0.082	0.081	ψ_5	0.259	0.123	0.322	0.107
β_6	0.004	0.040	0.005	0.052	ψ_6	0.035	0.058	0.053	0.067
β_7	0.118	0.079	0.130	0.079	ψ_7	0.043	0.072	0.281	0.113
β_8	0.747	0.073	0.092	0.077	ψ_8	0.384	0.132	0.188	0.118
β_η	-0.059	0.112	-0.039	0.092	ψ_η	1.205	0.106	1.910	0.104
σ^2	0.312	0.150	0.300	0.152					
λ_{21}	-0.791	0.062	-0.714	0.057					
λ_{31}	-0.865	0.067	-0.625	0.068					

To see more clearly, Table 4 gives the resulting selected variables according to SS and BaLsso. It can be seen that (i) for part one, both methods give the same results for the selection of factors ‘gender’, ‘age’, ‘material status’, ‘employment’, ‘number of adults’ and ‘family culture’. Two methods favor that ‘age’, ‘material status’, and ‘number of adults’ can be helpful in improving model fits while ‘gender’ and ‘family culture’ have less impacts on the probabilities of being held finance debt. However, there exist contradictory conclusion in selecting ‘health condition’, ‘education’ and ‘income’; (ii) for part two, except the factors ‘age’ and ‘number of adults’, two methods give the same results. In particular, both methods support that ‘family culture’ is relevant to the amount of household finance debts being held. This fact is also revealed by [17] in the analysis of CHFS by using two-part nonlinear latent variable model. The further interpretation is omitted for saving spaces.

Table 6. The selected variables in the CHFS data: 0: exclude and 1: included.

VAR	Part one		Part two	
	SS	BaLsso	SS	BaLsso
Gender	0	0	1	1
Age	1	1	1	0
Material status	1	1	0	0
Health condition	1	0	0	0
Education	1	0	1	1
Employment	0	0	0	0
No. of Adults	1	1	0	1
Income	1	0	1	1
Family culture	0	0	1	1

6. Discussion

Two-part latent variable model can be considered as an extension of traditional two-part model to the situations where the latent variables are included to identify the unobserved heterogeneity of population resulting from the absence of the observed covariates. When analyzing such a model, an important issue is to determine which factor is relevant to the outcome variable. This is especially true when the number of exogenous factors is high because the usual model selection/comparison procedure is extremely time-consuming. In this paper, we restor to the Bayesian variable selection method and developed a fully Bayesian variable selection procedure for the semi-continuous data.

Our formulation is along the lines with the spike and slab bimodal prior and recast the distribution of regression coefficients and factor loadings as hierarchy of priors over the parameter and model space. The selected variables is identified with high posterior probability of occurrence. We also consider a adaptive Bayesian lasso (BaLsso) for reference. To facilitate the computation, we recast the logistic regression model in part one as the flavor of normal mixture model by introducing latent Polya-gamma variables. which admits the conjugate conjugate full-conditional distributions for all regression coefficients, factor loadings and factor variables.

Although the Bayesian variable selection has its unique advantage, there are still some limitations that need to be considered with care. First, its computational complexity is high. Bayes SSL requires Monte Carlo sampling to estimate the posterior distribution, which can lead to slower calculation speed, especially when working with high-dimensional data sets. Secondly, the method is sensitive to hyperparameter and data distribution assumptions. The selection of the hyperparameters of the prior distribution, such as the ratio of spike to slab, lasso penalty parameters, and data distribution assumptions, will have a great impact on the results. When the data does not conform to the model convention, the performance of the model is poor. Therefore, these issues need to be carefully considered in practical application to ensure that the Bayesian SS method can be effectively applied to specific data sets.

The existing applications of the proposed methodology can be applied to more general latent variable models that include the the multilevel SEMs [50] and longitudinal dynamic variable models [16,51] with discrete variables. These extensions are left for further study.

Funding: This research was funded by National Nature Science Foundation of China (NNSF 11471161) and Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (15KJB110010).

Acknowledgments: We are thankful to Professor Xin-Yuan Song, Department of Statistics, The Chinese University of Hong Kong, Hong Kong, for providing us CHFS data.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

TPM	Two-part model
TPLVM	Two-part latent variable model
SS	Spike and slab bimodal prior
BaLsso	Bayesian lasso
MCMC	Markov Chains Monte Carlo
CHFS	China household finance survey

Appendix A

In this section, we will present some technical details on the full conditionals in the MCMC sampling. For ease of exposition, for any scalar or vector x , we use $p(x|\cdots)$ to denote the conditional distribution of x given ' \cdots '. Note that under the scenarios SS and BaLsso, the full conditionals of Ω , \mathbf{U}^* , \mathbf{Y}^* and θ are exactly the same. The following derivations are mainly based on the Bayes theorem.

1. Full conditional of $p(\Omega|\cdots)$

It follows from (8), (2) and (5) that

$$p(\Omega|\cdots) = \prod_{i=1}^n p(\omega_i|\cdots),$$

where

$$p(\omega_i|\cdots) \propto p(u_i, u_i^*|\omega_i, \alpha, \beta) p(z_i|u_i, \omega_i, \gamma, \psi, \sigma^2) p(\mathbf{y}_i^*|\omega_i, \mu, \Lambda) p(\omega_i|\Phi).$$

Let $\kappa_i^* = \kappa_i - u_i^*(\gamma + \mathbf{x}_i^T \boldsymbol{\beta}_x)$, $z_i^* = z_i - \gamma - \mathbf{x}_i^T \boldsymbol{\psi}_x$. By some algebra, it can be shown that

$$p(\omega_i | \dots) \stackrel{D}{=} N_m(\hat{\boldsymbol{\mu}}_{\omega_i}, \hat{\boldsymbol{\Sigma}}_{\omega_i}), \quad (\text{A1})$$

where

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{\omega_i} &= \hat{\boldsymbol{\Sigma}}_{\omega_i} \left[\boldsymbol{\beta}_{\omega} \kappa_i^* + \boldsymbol{\psi}_{\omega} u_i z_i^* / \sigma^2 + \boldsymbol{\Lambda}^T (\mathbf{y}_i^* - \boldsymbol{\mu}) \right], \\ \hat{\boldsymbol{\Sigma}}_{\omega_i} &= \left[\boldsymbol{\beta}_{\omega} \boldsymbol{\beta}_{\omega}^T u_i^* + \boldsymbol{\psi}_{\omega} \boldsymbol{\psi}_{\omega}^T u_i / \sigma^2 + \boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \boldsymbol{\Phi}^{-1} \right]^{-1}. \end{aligned}$$

Hence, draw of $\boldsymbol{\Omega}$ can be obtained by simulating ω_i independently from the normal distribution (A1).

2. Full conditional of $p(\mathbf{U}^* | \dots)$

Following the similar derivation in [41], it can be shown that given \mathbf{U} , $\boldsymbol{\Omega}$ and $\boldsymbol{\theta}$, \mathbf{U}^* is the entilted Polya-Gamma distribution given by

$$p(\mathbf{U}^* | \dots) = \prod_{i=1}^n \text{PG}(u_i^* | 1, \eta_i) \quad (\text{A2})$$

where $\eta_i = \alpha + \boldsymbol{\beta}^T \mathbf{w}_i$. Drawing u_i from this distribution can be achieved via rejection sampling, see [41] or [52] for more details on this issue.

3. Full of conditional of $p(\mathbf{Y}^* | \dots)$

Note that

$$\begin{aligned} p(\mathbf{Y}^* | \dots) &\propto p(\mathbf{Y} | \mathbf{Y}^*, \boldsymbol{\delta}) p(\mathbf{Y}^* | \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \\ &= \prod_{i=1}^n \prod_{k=1}^p \left(\sum_{\ell=0}^c I\{y_{ik} = \ell, \delta_{k\ell} < y_{ik}^* \leq \delta_{k\ell+1}\} \right) \times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (y_{ik}^* - \mu_k - \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i)^2 \right\}. \end{aligned}$$

Hence, given $\boldsymbol{\Omega}$, the full conditional of \mathbf{Y}^* only depends on $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, \mathbf{Y} and $\boldsymbol{\Omega}$, and is given by

$$\begin{aligned} p(\mathbf{Y}^* | \dots) &= \prod_{i=1}^n \prod_{k=1}^p p(y_{ik}^* | \omega_i, \boldsymbol{\theta}, y_{ik}), \\ p(y_{ik}^* | \omega_i, \boldsymbol{\theta}, y_{ik}) &= N(\mu_k + \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i, 1) I\{\delta_{k,y_{ik}} < y_{ik}^* \leq \delta_{k,y_{ik}+1}\}. \end{aligned} \quad (\text{A3})$$

This is the truncated normal distribution and its draw can be obtained via inverse distribution sampling method, see for example, [53].

4. Full conditional of $p(\boldsymbol{\theta} | \dots)$

Recall that $\boldsymbol{\theta}$ is consisted of α , $\boldsymbol{\beta}$, γ , $\boldsymbol{\psi}$, σ^2 , $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\delta}$. Hence, draw of $\boldsymbol{\theta}$ can be accomplished by (i) drawing α from $p(\alpha | \dots)$, (ii) drawing $\boldsymbol{\beta}$ from $p(\boldsymbol{\beta} | \dots)$, (iii) drawing γ from $p(\gamma | \dots)$, (iv) drawing $(\boldsymbol{\psi}, \sigma^2)$ from $p(\boldsymbol{\psi}, \sigma^2 | \dots)$, (v) drawing $\boldsymbol{\mu}$ from $p(\boldsymbol{\mu} | \dots)$, (vi) drawing $\boldsymbol{\Lambda}$ from $p(\boldsymbol{\Lambda} | \dots)$, (vii) drawing $\boldsymbol{\Phi}$ from $p(\boldsymbol{\Phi} | \dots)$, and (viii) drawing $\boldsymbol{\delta}$ from $p(\boldsymbol{\delta} | \dots)$ sequentially. Note that given \mathbf{U}^* , \mathbf{Y}^* and $\boldsymbol{\Omega}$, the models (8), (2) and (5) reduce to the ordinary regression models, and hence most of full conditionals, similar to that of the regression coefficients and variance/covariance in the Bayesian regression analysis, are the standard distributions such as normal, gamma, inverse gamma and wishart distributions. As a matter of fact, by some tedious but non-trivial calculations, it can be shown that

$$p(\alpha | \dots) = N(\hat{\mu}_{\beta}, \hat{\sigma}_{\beta}^2), \quad p(\boldsymbol{\beta} | \dots) = N_q(\hat{\boldsymbol{\mu}}_{\beta}, \hat{\boldsymbol{\Sigma}}_{\beta}), \quad (\text{A4})$$

$$p(\gamma | \dots) = N(\hat{\mu}_{\gamma}, \hat{\sigma}_{\gamma}^2), \quad p(\boldsymbol{\psi}, \sigma^2 | \dots) = IG(\hat{\alpha}_{\sigma}, \hat{\beta}_{\sigma}) \times N_q(\hat{\boldsymbol{\mu}}_{\psi}, \sigma^2 \hat{\boldsymbol{\Sigma}}_{\psi}), \quad (\text{A5})$$

$$p(\boldsymbol{\mu} | \boldsymbol{\Omega}, \boldsymbol{\Lambda}, \mathbf{Y}^*) = N_p(\hat{\mathbf{m}}_{\mu}, \hat{\boldsymbol{\Sigma}}_{\mu}), \quad p(\boldsymbol{\Lambda} | \dots) = \prod_{k=1}^p p(\boldsymbol{\Lambda}_k | \dots) = \prod_{k=1}^p N_m(\hat{\boldsymbol{\Lambda}}_k, \hat{\mathbf{H}}_k), \quad (\text{A6})$$

$$p(\boldsymbol{\Phi}^{-1} | \dots) = W_m(\rho + n, \hat{\mathbf{R}}), \quad (\text{A7})$$

in which

$$\begin{aligned}
\hat{\mu}_\alpha &= \hat{\sigma}_\alpha^2 \sum_{i=1}^n (\kappa_i - u_i^* \beta^T \mathbf{w}_i), \quad \hat{\sigma}_\alpha^2 = \left(\sum_{i=1}^n u_i^* + \sigma_{\alpha 0}^{-2} \right)^{-1}, \\
\hat{\mu}_\beta &= \hat{\Sigma}_\beta \sum_{i=1}^n \mathbf{w}_i (\kappa_i - \alpha u_i^*), \quad \hat{\Sigma}_\beta^{-1} = \sum_{i=1}^n u_i^* \mathbf{w}_i \mathbf{w}_i + \text{diag}\{\gamma_\beta^{-2}\}, \\
\hat{\mu}_\gamma &= \hat{\sigma}_\gamma^2 \sum_{i=1}^n u_i (z_i - \psi^T \mathbf{w}_i) / \sigma^2, \quad \hat{\sigma}_\gamma^2 = \left(\sum_{i=1}^n u_i / \sigma^2 + \sigma_{\gamma 0}^{-2} \right)^{-1}, \\
\hat{\mu}_\psi &= \hat{\Sigma}_\psi \sum_{i=1}^n \mathbf{w}_i (z_i - \gamma) u_i / \sigma^2, \quad \hat{\Sigma}_\psi^{-1} = \sum_{i=1}^n u_i \mathbf{w}_i \mathbf{w}_i + \text{diag}\{\gamma_\psi^{-2}\}, \\
\hat{\alpha}_\sigma &= a_0 + |\mathcal{I}|/2, \\
\hat{\beta}_\sigma &= b_0 + \frac{1}{2} \left(\sum_{i=1}^n u_i z_i^2 - \hat{\mu}_\psi^T \hat{\Sigma}_\psi^{-1} \hat{\mu}_\psi + \Lambda_{0k}^T \mathbf{H}_{0k}^{-1} \Lambda_{0k} \right), \\
\hat{\mathbf{m}}_\mu &= \hat{\Sigma}_\mu (\Sigma_0^{-1} \mu_0 + n(\bar{\mathbf{Y}}^* - \Lambda \bar{\Omega})), \quad \hat{\Sigma}_\mu^{-1} = n\mathbf{I}_p + \Sigma_0^{-1}, \\
\hat{\Lambda}_k &= \hat{\mathbf{H}}_k (\mathbf{H}_{0k}^{-1} \Lambda_{0k} + \Omega^T \mathbf{Y}_{[k]}^{**}), \quad \hat{\mathbf{H}}_k^{-1} = n\Phi^{-1} + \Omega^T \Omega, \\
\hat{\mathbf{R}}^{-1} &= \mathbf{R}_0^{-1} + \Omega^T \Omega,
\end{aligned}$$

where \mathbf{Y}^{**} is the $n \times p$ matrix with the i^{th} row $\mathbf{y}_i^{*T} - \mu^T$, $\mathbf{Y}_{[k]}^{**}$ is the k^{th} column of \mathbf{Y}^{**} , and Ω is the $n \times m$ matrix with the i^{th} row ω_i ; $\bar{\mathbf{Y}}^* = \sum_{i=1}^n \mathbf{y}_i^* / n$, $\bar{\Omega} = \sum_{i=1}^n \omega_i / n$. are the sample means of \mathbf{Y}^* and Ω and $|\mathcal{I}|$ denotes the size of $\mathcal{I} = \{u_i = 1\}$.

However, for δ , we note that

$$\begin{aligned}
p(\delta | \dots) &= \prod_{k=1}^p p(\delta_k | \mathbf{Y}_{[k]}^*, \mathbf{Y}_{[k]}), \quad \text{and} \\
p(\delta_k | \mathbf{Y}_{[k]}^*, \mathbf{Y}_{[k]}) &\propto p(\delta_k) \prod_{i=1}^n \prod_{\ell=0}^c I\{y_{ik} = \ell, \delta_{k\ell} < y_{ik}^* \leq \delta_{k,\ell+1}\}.
\end{aligned}$$

Hence, drawing δ can be obtained by drawing δ_k from $p(\delta_k | \dots)$ independently. Moreover, under prior (24), it can be shown that

$$p(\delta_{k\ell} | \delta_{k,(-\ell)}, \mathbf{Y}_{[k]}^*, \mathbf{Y}_{[k]}) \propto p(\delta_{k\ell}, \delta_{k,(-\ell)}) I\left\{ \max_{y_{ik}=\ell-1} \{y_{ik}^*\}, \leq \delta_{k\ell} < \min_{y_{ik}=\ell} \{y_{ik}^*\} \right\}.$$

where $\delta_{k,(-\ell)}$ is the vector of δ_k with $\delta_{k\ell}$ removed. Let $h_{k,\ell} = \max\{\delta_{k,\ell-1}, \max_{y_{ik}=\ell-1} \{y_{ik}^*\}\}$, $g_{k,\ell} = \min\{\delta_{k,\ell+1}, \min_{y_{ik}=\ell} \{y_{ik}^*\}\}$. It follows from (25) that

$$\frac{F_0(\delta_{k\ell}) - F_0(\delta_{k,\ell-1})}{F_0(\delta_{k,\ell+1}) - F_0(\delta_{k,\ell-1})} | \delta_{k,-\ell}, \mathbf{Y}_{[k]}^*, \mathbf{Y}_{[k]} \sim \text{Beta}(\eta_{k,\ell}, \eta_{k,\ell+1}) I\{(s_{k,\ell}, t_{k,\ell})\}, \quad (\text{A8})$$

where

$$s_{k,\ell} = \frac{F_0(h_{k,\ell}) - F_0(\delta_{k,\ell-1})}{F_0(\delta_{k,\ell+1}) - F_0(\delta_{k,\ell-1})}, \quad t_{k,\ell} = \frac{F_0(g_{k,\ell}) - F_0(\delta_{k,\ell-1})}{F_0(\delta_{k,\ell+1}) - F_0(\delta_{k,\ell-1})}.$$

As a result, we can draw $\delta_{k\ell}$ by first generating a $\delta_{k\ell}^*$ from the truncated beta distribution (A8) and then transform it to the $\delta_{k\ell}$ via inverse-transformation by setting $F_0^{-1}(\delta_{k\ell}^* [F_0(\delta_{k,\ell+1}) - F_0(\delta_{k,\ell-1})] + F_0(\delta_{k,\ell-1}))$. A draw of truncated beta distribution can be obtained by implementing inverse-distribution sampling method.

4. Full conditional of $p(\mathbf{Q}^* | \dots)$

First of all, it is noted that \mathbf{Q}^* is consisted of \mathbf{F}_β , \mathbf{F}_ψ , w_β , w_ψ , η_β^{-2} , and η_ψ^{-2} under SS, and formed by γ_β^2 , γ_ψ^2 , λ_β^2 and λ_ψ^2 under BaLsso. Similar to that of θ , we update \mathbf{Q}^* by drawing observations from their full conditionals per component sequentially.

Firstly, it is noted that

$$p(\mathbf{F}_\beta | \dots) \propto \prod_{k=1}^q p(\beta_k | f_{\beta k}, \eta_{\beta k}^2) p(f_{\beta k} | w_\beta),$$

$$p(\mathbf{F}_\psi | \dots) \propto \prod_{k=1}^q p(\psi_k | \sigma^2, f_{\psi k}, \eta_{\psi k}^2) p(f_{\psi k} | w_\psi),$$

which indicates that the components in the posteriors of \mathbf{F}_β and \mathbf{F}_ψ are independent. Further, it follows easily from (12) that

$$p(f_{\beta k} | w_\beta, \eta_{\beta k}^2, \beta_k) = (1 - \hat{q}_{\beta k}) \delta_{\nu_{\beta 0}}(\cdot) + \hat{q}_{\beta k} \delta_1(\cdot),$$

$$p(f_{\psi k} | w_\psi, \eta_{\psi k}^2, \beta_k) = (1 - \hat{q}_{\psi k}) \delta_{\nu_{\psi 0}}(\cdot) + \hat{q}_{\psi k} \delta_1(\cdot),$$

where

$$\hat{q}_{\beta k} = \frac{w_\beta \phi(\beta_k / \eta_{\beta k})}{(1 - w_\beta) \phi(\beta_k / (\sqrt{\nu_{\beta 0}} \eta_{\beta k}) / \sqrt{\nu_{\beta 0}}) + w_\beta \phi(\beta_k / \eta_{\beta k})},$$

$$\hat{q}_{\psi k} = \frac{w_\psi \phi(\psi_k / (\sigma \eta_{\psi k}))}{(1 - w_\psi) \phi(\psi_k / (\sigma \sqrt{\nu_{\psi 0}} \eta_{\psi k}) / \sqrt{\nu_{\psi 0}}) + w_\psi \phi(\psi_k / (\sigma \eta_{\psi k}))},$$

and $\phi(\cdot)$ is the standard normal probability density function.

Secondly, it is noted that

$$p(w_\beta | \mathbf{F}_\beta) \propto p(w_\beta) p(\mathbf{F}_\beta | w_\beta) = p(w_\beta) \prod_{k=1}^q p(f_{\beta k} | w_\beta)$$

$$= c w_\beta^{a_\beta - 1} (1 - w_\beta)^{b_\beta - 1} \prod_{k=1}^q w_\beta^{I\{f_{\beta k}=1\}} (1 - w_\beta)^{I\{f_{\beta k}=\nu_{\beta 0}\}},$$

$$p(w_\psi | \mathbf{F}_\psi) \propto p(w_\psi) p(\mathbf{F}_\psi | w_\psi) = p(w_\psi) \prod_{k=1}^q p(f_{\psi k} | w_\psi)$$

$$= c w_\psi^{a_\psi - 1} (1 - w_\psi)^{b_\psi - 1} \prod_{k=1}^q w_\psi^{I\{f_{\psi k}=1\}} (1 - w_\psi)^{I\{f_{\psi k}=\nu_{\psi 0}\}}.$$

Hence,

$$p(w_\beta | \dots) = \text{Beta}(c_{\beta 1} + |\{f_{\beta k} = 1\}|, c_{\beta 2} + |\{f_{\beta k} = \nu_{\beta 0}\}|), \quad (\text{A9})$$

$$p(w_\psi | \dots) = \text{Beta}(c_{\psi 1} + |\{f_{\psi k} = 1\}|, c_{\psi 2} + |\{f_{\psi k} = \nu_{\psi 0}\}|), \quad (\text{A10})$$

where $|A|$, as before, is the size of set A .

Lastly, it follows from

$$\begin{aligned} p(\boldsymbol{\eta}_{\beta}^{-2} | \mathbf{F}_{\beta}, \boldsymbol{\beta}) &\propto p(\boldsymbol{\beta} | \mathbf{F}_{\beta}, \boldsymbol{\eta}_{\beta}^{-2}) p(\boldsymbol{\eta}_{\beta}^{-2}) \\ &= \prod_{k=1}^q (\eta_{\beta k}^{-2})^{1/2} \exp \left\{ -\frac{1}{2} \eta_{\beta k}^{-2} \beta_k^2 / f_{\beta k} \right\} (\eta_{\beta k}^{-2})^{a_{\beta 1}-1} \exp \{ -a_{\beta 2} \eta_{\beta k}^{-2} \}, \\ p(\boldsymbol{\eta}_{\psi}^{-2} | \mathbf{F}_{\psi}, \boldsymbol{\psi}) &\propto p(\boldsymbol{\psi} | \mathbf{F}_{\psi}, \boldsymbol{\eta}_{\psi}^{-2}) p(\boldsymbol{\eta}_{\psi}^{-2}) \\ &= \prod_{k=1}^q (\eta_{\psi k}^{-2})^{1/2} \exp \left\{ -\frac{1}{2} \eta_{\psi k}^{-2} \psi_k^2 / f_{\psi k} \right\} (\eta_{\psi k}^{-2})^{a_{\psi 1}-1} \exp \{ -a_{\psi 2} \eta_{\psi k}^{-2} \} \end{aligned}$$

that

$$\begin{aligned} p(\boldsymbol{\eta}_{\beta}^{-2} | \mathbf{F}_{\beta}, \boldsymbol{\beta}) &= \prod_{k=1}^q p(\eta_{\beta k}^{-2} | f_{\beta k}, \beta_k) = \prod_{k=1}^q Ga(\hat{\tau}_{\beta k}, \hat{\zeta}_{\beta k}), \\ p(\boldsymbol{\eta}_{\psi}^{-2} | \mathbf{F}_{\psi}, \boldsymbol{\psi}) &= \prod_{k=1}^q p(\eta_{\psi k}^{-2} | f_{\psi k}, \psi_k) = \prod_{k=1}^q Ga(\hat{\tau}_{\psi k}, \hat{\zeta}_{\psi k}), \end{aligned}$$

where

$$\begin{aligned} \hat{\tau}_{\beta k} &= \tau_{\beta 0} + 1/2, \quad \hat{\zeta}_{\beta k} = \zeta_{\beta 0} + \beta_k^2 / (2f_{\beta k}), \\ \hat{\tau}_{\psi k} &= \tau_{\psi 0} + 1/2, \quad \hat{\zeta}_{\psi k} = \zeta_{\psi 0} + \psi_k^2 / (2f_{\psi k}). \end{aligned}$$

For BaLasso, we follow the practice in [37] and can show

$$\begin{aligned} p(\boldsymbol{\gamma}_{\beta}^{-2} | \cdots) &= \prod_{k=1}^q p(\gamma_{\beta k}^{-2} | \cdots) = \prod_{k=1}^q IG(\hat{\mu}_{\beta}, \hat{\lambda}_{\beta}), \\ p(\boldsymbol{\gamma}_{\psi}^{-2} | \cdots) &= \prod_{k=1}^q p(\gamma_{\psi k}^{-2} | \cdots) = \prod_{k=1}^q IG(\hat{\mu}_{\psi}, \hat{\lambda}_{\psi}), \end{aligned}$$

in which

$$\begin{aligned} \hat{\mu}_{\beta} &= \sqrt{\hat{\lambda}_{\beta} / \beta_j^2}, \quad \hat{\lambda}_{\beta} = \lambda_{\beta k}^2, \\ \hat{\mu}_{\psi} &= \sqrt{\sigma^2 \hat{\lambda}_{\psi} / \beta_j^2}, \quad \hat{\lambda}_{\psi} = \lambda_{\psi k}^2, \end{aligned}$$

where $IG(\mu, \lambda)$ is the inverse-gaussian distribution with density $\sqrt{\lambda/(2\pi)} x^{-3/2} \exp\{-\lambda(x - \mu)^2/(2\mu^2 x)\} (x > 0)$ [54].

Similarly,

$$\begin{aligned} p(\boldsymbol{\lambda}_{\beta}^2 | \cdots) &= \prod_{k=1}^q p(\lambda_{\beta k}^2 | \cdots) = \prod_{k=1}^q Ga(\hat{a}_{\beta k}, \hat{b}_{\beta k}), \\ p(\boldsymbol{\lambda}_{\psi}^2 | \cdots) &= \prod_{k=1}^q p(\lambda_{\psi k}^2 | \cdots) = \prod_{k=1}^q Ga(\hat{c}_{\psi k}, \hat{d}_{\psi k}), \end{aligned}$$

in which

$$\begin{aligned} \hat{a}_{\beta k} &= a_{k0} + 1.0, \quad \hat{b}_{\beta k} = b_{k0} + 0.5\gamma_{\beta k}^2, \\ \hat{c}_{\psi k} &= c_{k0} + 1.0, \quad \hat{d}_{\psi k} = d_{k0} + 0.5\gamma_{\psi k}^2. \end{aligned}$$

References

1. Deb, P, Munkin, M. K., Trivedic, R. K. (2006). Bayesian analysis of the two-part model with endogeneity: Application to health care expenditure. *J. Appl. Econ.*, **21**, 1081-1099.
2. Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, **39**, 829-844.
3. Neelon, B., Zhu, L. and Neelon, S. E. B. (2015). Bayesian two-part spatial models for semicontinuous data with application to emergency department expenditures. *Biostatistics*, **16**, 465-479.
4. Manning, W. G., et al. (1981). A two-part model of the demand for medical Care: preliminary results from the health insurance experiment, in *Health, Economics, and Health Economics*, eds. van der Gaag, J. and Perlman, M., p. 103-104, Amsterdam: North-Holland.
5. Su, L., Tom, B. D. and Farewell, V. T. (2009). Bias in 2-part mixed models for longitudinal semi-continuous data. *Biostatistics*, **10**, 374-389.
6. Su, L., Tom, B. D. and Farewell, V. T. (2015). A likelihood-based two-part marginal model for longitudinal semi-continuous data. *Statistical Methods in Medical Research*, **24**, 194-205.
7. Liu, L., Cowen, M. E., Strawderman, R. L. and Shih, Y. C. T. (2010). A flexible two-part random effects model for correlated medical costs. *Journal of Health Economics*, **29**, 110-123.
8. Smith, V. A., Neelon, B., Preisser, J. S., Maciejewski, L. (2015). A marginalized two-part model for semicontinuous data. *Statistics in Medicine*, **33**, 4891-4903.
9. Tooze, J. A., Grunwald, J. K. and Jones, R. H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research*, **11**, 341-355.
10. Brown, R. A., Monti, P. M., Myers, M. G., Martin, R. A., Rivinus, T., Dubreuil, M. E. T. and Rohsenow, D. J. (1998). Depression among cocaine abusers in treatment: Relation to cocaine and alcohol use and treatment outcome. *American Journal of Psychiatry*, **155**, 220-225.
11. Olsen, M. K. and Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data, *Journal of the American Statistical Association*, **96**, 730-745.
12. Xing, D. Y., Huang, Y. X., Chen H. N., Zhu, Y. L., Dagen, G. A. and Baldwin, J. (2017). Bayesian inference for two-part mixed effects model using skew distributions, with application to longitudinal semi-continuous alcohol data, *Statistical Methods in Medical Research*, **26**, 1838-1853.
13. Chen, J. Y., Zheng, L. Y. and Xia, Y. M. (2023). Bayesian analysis for two-part latent variable model with application to fractional data. *Communications in Statistics - Theory and Methods*, published online: DOI:10.1080/03610926.2023.2273205.
14. Kim, Y., and Muthén, B. O. (2009). Two-Part Factor Mixture Modeling: Application to an Aggressive Behavior Measurement Instrument. *Structural Equation Modeling: A Multidisciplinary Journal*, **16**, 602-624.
15. Feng, X., Lu, B., Song, X. and Ma, S. (2019). Financial literacy and household finances: A Bayesian two-part latent variable modeling approach. *Journal of Empirical Finance*, **51**, 119-137.
16. Xia, Y. M. and Tang, N. S. (2019). Bayesian analysis for mixture of latent variable hidden Markov models with multivariate longitudinal data. *Computational Statistics & Data Analysis*, **132**, 190-211.
17. Gou, J. W., Xia, Y. M. and Jiang, D. P. (2023). Bayesian analysis of two-part nonlinear latent variable model: Semiparametric method. *Statistical Modelling*, **23** (4), 721-741.
18. Xiong, S. C., Xia, Y. M., Lu, B. (2023). Bayesian Analysis of Two-Part Latent Variable Model with Mixed Data. *Communications in Mathematics and Statistics*, in press, DOI:10.1007/s40304-023-00359-1.
19. Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348-1360.
20. Fu, W. J. (1998). Penalized regression: the bridge versus the lasso. *Journal of computational and Graphical Statistics*, **7**, 109-148.
21. Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer-Verlag, New York, NY.
22. Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity - The Lasso and Generalization*. CRC Press: New York.
23. Kuo, L. and Mallick, B. K. (1998). Variable selection for regression models. *Sankhya, Ser. B*, **60**, 65-81.
24. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B*, **58**, 267-288.

25. Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B*, **67**, 301-320.
26. Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American statistical Association*, **101**, 1418-1429.
27. Zhang, W., T. Ota, V. Shridhar, J. Chien, B. Wu et al., (2013). Network-based survival analysis reveals subnetwork signatures for predicting outcomes of ovarian cancer treatment. *PLOS Comput. Biol.*, **9**, e1002975.
28. Zhao, Q., X. Shi, Y. Xie, J. Huang, B. Shia et al., (2014). Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Brief. Bioinform.*, **16**, 291-303.
29. George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**, 881-889.
30. George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Stat. Sin.*, **7**, 339-373.
31. Chipman, H. A. (1996). Bayesian variable selection with related predictors. *Canad. J. Statist.*, **24**, 17-36.
32. Ishwaran, H., and Rao, J. S. (2005). Spike and Slab gene selection for multigroup microarray data. *Journal of the American Statistical Association*, **87**, 371-390.
33. Ishwaran, H., and Rao, J. S. (2005). Spike and Slab variable selection: frequentist and Bayesian strategies. *The Annals of Statistics*, **33**, 730-773.
34. Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, **83**(404):1023-1032.
35. Rockova, V. and George, E. I. (2014). EMVS: The EM approach to Bayesian variable selection. *Journal of the American Statistical Association*, **109**, 828-846.
36. Tang, Z. X., Shen, Y. P., Xinyan Zhang, X. Y. and Nengjun Yi, N. J. (2017) The Spike-and-Slab Lasso Generalized Linear Models for Prediction and Associated Genes Detection. *Genetics*, **205**, 77-88.
37. Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, **103**(482), 681-686.
38. Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized latent variable modelling: multilevel, longitudinal and structural equation models*. Chapman & Hall/CRC: London.
39. Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons: New York.
40. Lee, S. Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. John Wiley & Sons: New York.
41. Polson, N. G., Scott, J. G., Windle, J. (2013). Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables. *Journal of the American Statistical Association*, **108**, 1339-1349.
42. Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd Edn. John Wiley & Sons: New York.
43. Sha, N. J. and Dechi, B. O. (2019). A Bayes inference for ordinal response with latent variable approach. *Stats*, **2**, 321-331;
44. Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American statistical Association*, **82**, 528-550.
45. Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.
46. Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6), 721-741.
47. Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**, 457-511.
48. Feng, X., Wang, Y. F., Lu, B., and Song, X. Y. (2017). Bayesian regularized quantile structural equation models. *Journal of Multivariate Analysis*, **154**, 234-248.
49. Little, R. J. A. and Rubin, D. B. (2002). *Statistical analysis with missing data*, second Edition. John Wiley & Sons: New York.
50. Song, X. Y. and Lee, S. Y. (2012). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology*, **56**(3), 135-148.
51. Song, X. Y., Xia, Y. M. and Zhu H. T. (2017). Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics*, **73**, 313-323.
52. Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag: New York.
53. Ross, S. M. (1991). *A Course in Simulation*. MacMillan: New York.

54. Chhikara, R. S., and Folks, L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*, New York: Marcel Dekker.
55. Duan, N., Manning, W. G., Morris, C. N. and Newhouse, J. P. (1983). A Comparison of alternative models for the demand for medical Care. *Journal of Business and Economic Statistics*, **1**, 115-126.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.