MDPI

Article

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

Qi Zhang [1], Yihui Zhang [2] and Yemao Xia [1,*]

1 School of Science, Nanjing Forestry University, Nanjing 210037, China; njfuzhangqi@njfu.edu.cn
2 School of Computer Science, China University of Geosciences, Wuhan 430074, China; yihuizhang@cug.edu.cn
* Correspondence: ymxia@njfu.edu.cn; Tel.: +86-25-8542-7703

Abstract: Semi-continuous data are very common in social sciences and economics. In this paper, a Bayesian variable selection procedure is developed to assess the influence of observed and/or unobserved exogenous factors on semi-continuous data. Our formulation is based on a two-part latent variable model with polytomous responses. We consider two schemes for the penalties of regression coefficients and factor loadings: a Bayesian spike and slab bimodal prior and a Bayesian lasso prior. Within the Bayesian framework, we implement a Markov chain Monte Carlo sampling method to conduct posterior inference. To facilitate posterior sampling, we recast the logistic model from Part One as a norm-type mixture model. A Gibbs sampler is designed to draw observations from the posterior. Our empirical results show that with suitable values of hyperparameters, the spike and slab bimodal method slightly outperforms Bayesian lasso in the current analysis. Finally, a real example related to the Chinese Household Financial Survey is analyzed to illustrate application of the methodology.

1. Introduction

Semi-continuous data, which are characterized by excessive zeros, are very common in the fields of social sciences and economics. A typical example is given by [1] in an 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 among users. For understanding such a type of data structure, a two-part model [2] is a widely appreciated statistical method. The basic assumption for a two-part model is that the overall model consists of two processes: one binary process (Part One) and one continuous positive-valued process (Part Two). The binary process, usually formulated within a logistic or probit regression model, is used to indicate whether the items have been responded to or not, while the continuous process, conditioning of the binary process, is used to describe what the actual levels of the responses are (see, e.g., [3]). By combining two processes into one, a two-part model provides a unified and flexible way to describe various relationships underlying semi-continuous data. Now, two-part models have been widely used for health services [4–6], medical expenditures [1,7–10], household finances [11], substance use studies [12,13], and genome analysis [14].

A traditional two-part model usually formulates exogenous explanatory factors as fixed and observed. However, in many real applications, especially for socials survey, many unobserved/latent and random factors also have important impacts on the outcome variable(s). This fact is revealed by [15] in a study of children's aggressive behavior. Ref. [15] noted that two factors, the propensity to engage in aggressive behavior and the

propensity to have highly aggressive activity levels, had significant influence on children's aggressive behavior. The authors incorporated two such latent factors into their analysis and established a two-component–two-part mixture model to identify the heterogeneity of the population. Ref. [16] noticed that in China, the financial literacy of a family had a non-ignorable effect on the desire to hold finance debts and also affected the amount of finance debt being held. They suggested conducting a joint analysis of latent factors and observed covariates in a two-part regression model. Latent factors are further manifested by multiple binary measurements via a factor analysis model. Ref. [17] incorporated a two-part regression model into a general latent variable model framework and analyzed the internal relationships between multiple factors longitudinally. These methods have brought significant attention to two-part models in behavioral science, economics, psychology, and medicine in recent years: see, for example [14,18,19], and references therein for further developments of two-part models.

In an analysis of semi-continuous data, an important issue is to determine which explanatory factors are helpful for 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, lassos and their extensions [20–27] have been the most commonly used methods for 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, [28,29]). Though more appealing, lasso-type regularization methods also suffer some limitations. For example, most contributions are developed within the frequentist framework, and their performance heavily depends on the large sample theory (see, for example [20,21,26,27], and references therein). This also readily leads to computational difficulty in the analysis of mixed data. An alternative to variable selection is conducted within the Bayesian framework. Statisticians have introduced hierarchical models with mixed spike-and-slab priors that can adaptively determine the amount of shrinkage [30,31]. The spike and slab prior is the fundamental basis for most Bayesian variable selection approaches and has proved remarkably successful [30–36]. Recently, Bayesian spike and slab priors have been applied to predictive modeling and variable selection in large-scale genomic studies: see [37] for a simple review. Nevertheless, model selection has never been considered in a two-part regression model with latent variables. In this study, we introduce a spike and slab model and Bayesian lasso that have been combined into a two-part latent variable model, which is a first attempt for this model.

Our formulation is more along the lines of the spike and slab bimodal prior in [34] and the Bayesian lasso in [38]. 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 a two-point mixture model with one component located at a point close to zero and the other component situated at a point far away zero. The mixing proportion is governed by a beta-distribution with suitable hyperparameters. The other scheme uses a Bayesian lasso for which the variance is specified via a gamma distribution that is scaled by the penalty parameters. The 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 chain Monte Carlo sampling method. A Gibbs sampler is used to draw observations from the posterior. We obtain all full conditionals. Posterior analysis is carried out based on the simulated observations. We investigate the performance of the proposed methods via a simulation study and a real example. Our empirical results show that the two schemes result in similar results for variable selection, but the spike and slab bimodal prior with suitable hyperparameters slightly outperforms the Bayesian lasso in terms of the correct rate.

The remainder of this paper is organized as follows. Section 2 introduces the proposed model for semi-continuous data with latent variables. Section 3 develops an MCMC sampling algorithm for the proposed model. Bayesian inference procedures to include parameter estimation and model assessment are also presented in this section. In Section 4, we present the results of a simulation study to assess the performance of the proposed methodology and illustrate the practical value of our proposed model by analyzing household finance debt data. Section 5 concludes the paper with a discussion. Some technical details are given in Appendix A.

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 feature extraction.

2.1. Two-Part Latent Variable Model

Suppose that for $i = 1, \ldots, n$, s_i is a semicontinuous outcome variable that takes a value in $[0, \infty)$; $\mathbf{x}_i$ is a generic vector with 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 $j = 1, \ldots, r$. Moreover, we include m latent/unobserved variables $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{im})^T$ into the analysis to account for the 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 with the noise. Inclusion of latent variables can improve model fit and strengthen the power of model interpretation: see [39] for more discussion of the latent variables in a general setting. To deal with the spike of s_i at zero, we follow the common routine in the literature (see, for example [10,12]) and identify s_i with two surrogate variables: $u_i = I\{s_i > 0\}$ and $z_i = \log(s_i | s_i > 0)$, where $I(A)$ denotes the indicator function of set A. That is, we separate the whole dataset into two parts: one part is the binary dataset that corresponds to the response-to-nonresponse indicators of the subject and the other part is the set of logarithms of positive values. Our interest focuses on the exploration of the effects of exogenous factors on the two parts.

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

$$\log\left(\frac{P(u_i = 1)}{1 - P(u_i = 1)}\right) = \eta_i^u = \alpha + \boldsymbol{\beta}_x^T \mathbf{x}_i + \boldsymbol{\beta}_\omega^T \boldsymbol{\omega}_i, \tag{1}$$

$$p(z_i | u_i = 1, \boldsymbol{\omega}_i) = N(\eta_i^z, \sigma^2), \tag{2}$$

$$\eta_i^z = \gamma + \boldsymbol{\psi}_x^T \mathbf{x}_i + \boldsymbol{\psi}_\omega^T \boldsymbol{\omega}_i,$$

in which α and γ are the scalars of the intercept parameters, $\boldsymbol{\beta}_x$ and $\boldsymbol{\psi}_x$ are the vectors of the regression coefficients, and $\boldsymbol{\beta}_\omega$ and $\boldsymbol{\psi}_\omega$ are the vectors of the factor loadings; σ^2 is the scale and 'T' is the transpose operator of the 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. Note that Equation (1) can be represented as

$$p(u_i | \mathbf{x}_i, \boldsymbol{\omega}_i) = \exp(u_i \eta_i^u) / (1 + \exp(\eta_i^u)), \ u_i = 0, 1, \tag{3}$$

and we refer to it as the logistic model.

The involvement of latent variables apparently complicates the model. It readily leads to model identification problems [40,41]. 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 a latent variable (LV) [40,41] approach. A basic assumption of the LV approach is that there exists, say, p manifestations $\mathbf{y}_i = (y_{i1}, \cdots, y_{ip})^T$,

of which each y_{ij} may be continuous, counted, or categorical, and we assume that they satisfy the following link equation:

$$F(\mathbf{y}_i, \boldsymbol{\omega}_i, \boldsymbol{\epsilon}_i, \boldsymbol{\varphi}) = \mathbf{0}, \tag{4}$$

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

$$y_{ij} = \ell_j \quad \text{if} \ \delta_{j,\ell_j} < y_{ij}^* \leq \delta_{j,\ell_j+1}, \tag{5}$$

where $\delta_{j,0} < \delta_{j,1} < \cdots < \delta_{j,c_j} < \delta_{j,c_j+1}$ are the threshold parameters satisfying $\delta_{j,0} = -\infty$ and $\delta_{j,c_j+1} = +\infty$, and $\mathbf{y}_i^* = (y_{i1}^*, \cdots, y_{ip}^*)^T$ is the vector of latent responses satisfying the factor analysis model:

$$\mathbf{y}_i^* = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \tag{6}$$

$$\boldsymbol{\omega}_i \overset{iid.}{\sim} N_m(\mathbf{0}, \boldsymbol{\Phi}), \ \boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \mathbf{I}_p), \ \text{and} \ \boldsymbol{\omega}_i \perp \boldsymbol{\epsilon}_i, \tag{7}$$

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

We refer to the model specified by (1), (2), and (5) associated with (6) 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 sharing 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 a traditional two-part model and ordinal regression model.

To facilitate efficient calculation and motivated by the key identity in [42] (see Equation (2) in their seminal paper), we express model (3) as a mixture model with form

$$\frac{\exp(u_i(\alpha + \boldsymbol{\beta}^T \mathbf{w}_i))}{1 + \exp(\alpha + \boldsymbol{\beta}^T \mathbf{w}_i)}$$
$$= 2^{-1} \exp(\kappa_i(\alpha + \boldsymbol{\beta}^T \mathbf{w}_i)) \int_0^\infty \exp\left\{-\frac{u_i^*}{2}(\alpha + \boldsymbol{\beta}^T \mathbf{w}_i)^2\right\} p_{PG}(u_i^*) \mathrm{d}u_i^*, \tag{8}$$

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 (3) can be considered to be the marginal density of the joint distribution

$$p(u_i, u_i^* \mid \mathbf{x}_i, \boldsymbol{\omega}_i) = 2^{-1} \exp\left\{\kappa_i \eta_i^u - \frac{u_i^*}{2} \eta_i^{u2}\right\} p_{PG}(u_i^*). \tag{9}$$

Note that the exponential part is the kernel of the 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 $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^n$ be the sets of observed variables. We write $\mathbf{\Omega} = \{\boldsymbol{\omega}_i\}_{i=1}^n$ for the collection of factor variables and $\mathbf{U}^* = \{u_i^*\}_{i=1}^n$, $\mathbf{Y}^* = \{\mathbf{y}_i^*\}_{i=1}^n$ for the sets of latent response variables. The complete data likelihood is given by

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

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

2.2. Bayesian Feature Selection

Generally speaking, regression variables $\mathbf{x}_i$ and factor variables $\boldsymbol{\omega}_i$ may not impact 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 far from zero. In the context of frequentist inference, this issue is generally tackled via stepwise regression, during which for each variable it is decided to be excluded 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 [38] and assume

$$
\boldsymbol{\beta} \sim N_q(\mathbf{0}, diag\{\gamma_{\beta k}^2\}), \quad \boldsymbol{\psi} \sim N_q(\mathbf{0}, \sigma^2 diag\{\gamma_{\psi k}^2\}), \tag{10}
$$

in which we use $diag\{a_k\}$ to represent a diagonal matrix with the kth diagonal element a_k and let $q = r + m$. That is, we assume that each β_k in $\boldsymbol{\beta}$ (ψ_k is similar) is centered at zero (or equivalently, each w_{ik} is excluded from $\mathbf{w}_i$), but the probability is 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 about $\gamma_{\beta k}^2$ and $\gamma_{\psi k}^2$ is that:

$$
\gamma_{\beta k}^2 \overset{ind.}{\sim} (1 - w_\beta) \delta_{\nu_{\beta 0} \eta_{\beta k}^2}(\cdot) + w_\beta \delta_{\eta_{\beta k}^2}(\cdot), \tag{11}
$$

$$
\gamma_{\psi k}^2 \overset{ind.}{\sim} (1 - w_\psi) \delta_{\nu_{\psi 0} \eta_{\psi k}^2}(\cdot) + w_\psi \delta_{\eta_{\psi k}^2}(\cdot), \tag{12}
$$

where $\delta_a(\cdot)$ is the Dirac measure concentrated at point a, w_β is the random weight used to measure the degree of 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 the slab; $\nu_{\beta 0}$ is a previously specified

small positive value used to identity the 'spike' of β_k at zero. In other words, every $\gamma^2_{\beta k}$ is assumed to be equal to $\eta^2_{\beta k}$ with probability w_β and equal to $\nu_{\beta 0} \eta^2_{\beta_k}$ 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:

$$p(w_\beta | a_\beta, b_\beta) = Beta(a_\beta, b_\beta), \quad p(w_\psi | a_\psi, b_\psi) = Beta(a_\psi, b_\psi), \tag{13}$$

where a_β, a_ψ, b_β, and b_ψ are the hyperparameters used to control the shape of the beta density: that is, to determine the magnitude of weights in $(0,1)$. For example, if $a_{\beta 1}$ in Equation (13) is small and $b_{\beta 1}$ is large, then Equation (13) encourages w_β to take a small value with high probability. In contrast, it follows from $1 - Beta(a_\beta, b_\beta) = Beta(b_\beta, a_\beta)$ that a large a_β and small b_β encourage w_β to take a large value in $(0,1)$. In the case that $a_\beta = b_\beta = 1.0$, Equation (13) reduces to a uniform distribution on $(0,1)$. In this case, every value in $(0,1)$ is possible for w_β with identical probability. In real applications, if no information is available, one can assign these values in a manner that ensures the beta distribution is inflated enough.

Finally, to measure the magnitudes of the 'slab' in the distributions of β_k and ψ_k, we specify gamma distributions for $\eta^{-2}_{\beta k}$ and $\eta^{-2}_{\psi k}$, or equivalently,

$$\eta^2_{\beta k} | a_{\beta 1}, a_{\beta 2} \stackrel{iid.}{\sim} IG(\tau_{\beta 0}, \zeta_{\beta 0}), \quad \eta^2_{\psi k} | a_{\psi 1}, a_{\psi 2} \stackrel{iid.}{\sim} IG(\tau_{\psi 0}, \zeta_{\psi 0}), \tag{14}$$

where we use '$IG(\tau, \zeta)$' to signify the inverse gamma distribution with mean $\zeta/(\tau - 1)$ for $\tau > 1$ and variance $\zeta^2/((\tau - 1)^2(\tau - 2))$ for $\tau > 2$; $\tau_{\beta 0}$, $\zeta_{\beta 0}$, $\tau_{\psi 0}$, and $\zeta_{\psi 0}$ are the hyperparameters, which are treated as fixed and known. Similarly, one can assign values to them to ensure that (14) is dispersed enough. For example, we can follow the routines in [34] for 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 (11) and (12) can be formulated as a hierarchy as follows: for $k = 1, \dots, q$,

$$\gamma^2_{\beta_k} = f_{\beta k} \eta^2_{\beta k}, \quad \gamma^2_{\psi k} = f_{\psi k} \eta^2_{\beta k},$$

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

$$f_{\psi k} | \nu_{\psi 0}, w_\psi \stackrel{iid.}{\sim} (1 - w_\psi) \delta_{\nu_{\psi 0}}(\cdot) + w_\psi \delta_1(\cdot), \tag{16}$$

where $f_{\beta k}$ and $f_{\psi k}$ are the latent binary variables. Such a formulation aims to separate $\eta^2_{\beta k}$ and $\eta^2_{\psi k}$ from the distributions (11) and (12) to facilitate posterior sampling.

It is instructive to compare the proposed method to the Bayesian lasso [38], in which the variance parameters $\gamma^2_{\beta k}$ and $\gamma^2_{\psi k}$ in Equation (10) are specified via exponential distributions as follows:

$$p(\gamma^2_\beta | \lambda^2_\beta) = \prod_{k=1}^{q} \frac{\lambda^2_{\beta k}}{2} \exp(-\lambda^2_{\beta k} \gamma^2_{\beta k}/2), \tag{17}$$

$$p(\gamma^2_\psi | \lambda^2_\psi) = \prod_{k=1}^{q} \frac{\lambda^2_{\psi k}}{2} \exp(-\lambda^2_{\psi k} \gamma^2_{\psi k}/2), \tag{18}$$

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

Modeling $\gamma^2_{\beta k}$ and $\gamma^2_{\psi k}$ like Equations (17) and (18) leads to marginal distributions of β_k and ψ_k as Laplace distributions with location zero and scale λ_k. The penalty parameters $\lambda^2_{\beta k}$ and $\lambda^2_{\psi k}$ are rather crucial for 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 λ is, the more kurtosis the density has, indicating more penalties on the regression coefficient.
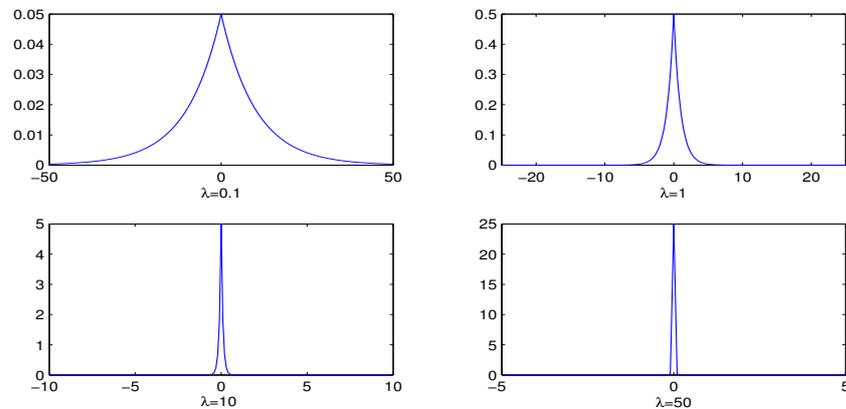
Figure 1. Plot of the densities of Laplace distribution for different choices of λ.

Due to the key roles in Equations (17) and (18) of λ_β^2 and λ_ψ^2, we assign the following gamma priors to them, i.e.,

$$p(\lambda_\beta^2) = \prod_{k=1}^{q} p(\lambda_{\beta k}^2) = \prod_{k=1}^{q} Ga(a_{k0}, b_{k0}), \tag{19}$$

$$p(\lambda_\psi^2) = \prod_{k=1}^{q} p(\lambda_{\psi k}^2) = \prod_{k=1}^{q} Ga(c_{k0}, d_{k0}), \tag{20}$$

where '$Ga(\nu, \lambda)$' denotes a gamma distribution with mean ν/λ. As previously discussed, 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 [43], one can set $a_{k0} = c_{k0} = 1$ and $b_{k0} = d_{k0} = 0.05$ 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 the known hyperparameters. Note that γ_β^2 and γ_ψ^2 are totally determined by $\mathbf{F}_\beta^*$, $\mathbf{F}_\psi^*$ and η_β^2, η_ψ^2. In the following, we abbreviate the spike and slab bimodal prior to SS and the 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 unknown parameters to complete Bayesian model specification. Based on the model convention, it is natural to assume that the parameters involved in the different models are independent.

Firstly, for $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Phi}$, we consider the following conjugate priors:

$$p(\boldsymbol{\mu}) = N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \tag{21}$$

$$p(\boldsymbol{\Lambda}) = \prod_{k=1}^{p} p(\boldsymbol{\Lambda}_k) = \prod_{k=1}^{p} N_m(\boldsymbol{\Lambda}_{0k}, \mathbf{H}_{0k}), \tag{22}$$

$$p(\boldsymbol{\Phi}) = IW(\rho_0, \mathbf{R}_0^{-1}), \tag{23}$$

where '$IW(\rho, \mathbf{R})$' denotes an inverse Wishart distribution with degrees of freedom ρ and scale matrix $\mathbf{R}$ [44]; $\boldsymbol{\Lambda}_k^T$ is the kth row vector of $\boldsymbol{\Lambda}$; $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0(p \times p) > 0$, $\boldsymbol{\Lambda}_{0k}$, $\mathbf{H}_{0k}(m \times m) > 0$, $\rho_0 > 0$, and $\mathbf{R}_0(m \times m) > 0$ are the hyperparameters, which are treated as fixed and known.

Secondly, for α, γ, and σ^2 in Parts One and Two, we assume they are mutually independent and satisfy

$$p(\alpha) = N(\alpha_0, \sigma_{\alpha 0}^2), \ \ p(\gamma) = N(\gamma_0, \sigma_{\gamma 0}^2), \ \ p(\sigma^{-2}) = Ga(a_0, b_0), \quad (24)$$

where α_0, $\sigma_{\alpha 0}^2$, γ_0, $\sigma_{\gamma 0}^2$, a_0, and 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 c. Moreover, we assume that $p(\delta) = \prod_{j=1}^p p(\delta_j)$, where $\delta_j = (\delta_{jk})$ is the jth 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 the distribution function of Student's t-distribution with degrees of freedom ν_0, where $\Phi(\cdot)$ is the standard normal distribution function. To specify a prior for δ, we follow [45] and let $p_j = F_0(\delta_j) - F_0(\delta_{j-1})$ for $j = 1, \cdots, c$. It is easily shown that this transformation is invertible with Jacobi determination unity. We first consider the following Dirichlet distribution for $\boldsymbol{p} = (p_1, \cdots, p_c)^T$:

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

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

$$\pi(\delta) = \frac{1}{\mathcal{B}(\eta_1, \dots, \eta_{c+1})} p_1^{\eta_1 - 1} \cdots p_c^{\eta_c - 1} (1 - \sum_{\ell=1}^c p_\ell)^{\eta_{c+1} - 1} \prod_{j=1}^c f_0(\delta_j) I\{\delta_1 < \cdots < \delta_c\}, \quad (25)$$

where $f_0(x)$ is the derivative of $F_0(x)$ with respect to x. We call (25) the transformed Dirichlet prior and use it as the prior of δ. An advantage of working with (25) is that, conditional upon δ_{j-1} and δ_{j+1}, the transformed distribution of δ_j has the beta distribution given by

$$\frac{F_0(\delta_j) - F_0(\delta_{j-1})}{F_0(\delta_{j+1}) - F_0(\delta_{j-1})} | (\delta_{j-1}, \delta_{j+1}) \sim Beta(\eta_j, \eta_{j+1}), (j = 1, \dots, c). \quad (26)$$

3.2. MCMC Sampling

With the prior given above, the inference about $\boldsymbol{\theta}$ is based on the posterior $p^o(\boldsymbol{\theta}|\mathbf{U}, \mathbf{Z}, \mathbf{Y})$, which has no closed form. Motivated by the key idea in [46], we treat latent quantities as the missing data and augment them to the observed data to form the complete data. The statistical inference is carried out based on the complete data likelihood. To this end, apart from $\boldsymbol{\Omega}$, $\mathbf{U}^*$, and $\mathbf{Y}^*$ mentioned before, we further let $\mathbf{Q}^*$ be the collection of latent quantities involved in the specifications of $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$, 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

$$p^{joint}(\boldsymbol{\Omega}, \mathbf{U}^*, \mathbf{Y}^*, \mathbf{Q}^*, \boldsymbol{\theta}|\mathbf{U}, \mathbf{Z}, \mathbf{Y}), \quad (27)$$

where p^o can be considered as the marginal of p^{joint}. We use the Markov chain Monte Carlo (MCMC) [47,48] sampling method to simulate observations from this target distribution. In particular, a Gibbs sampler is implemented to draw observations iteratively from the full conditional distributions as follows:

- Draw $\boldsymbol{\Omega}$ from $p(\boldsymbol{\Omega} \mid \mathbf{U}^*, \mathbf{Q}^*, \mathbf{Y}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{V})$;
- Draw $\mathbf{U}^*$ from $p(\mathbf{U}^* \mid \boldsymbol{\Omega}, \mathbf{Y}^*, \mathbf{Q}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{V})$;
- Draw $\mathbf{Y}^*$ from $p(\mathbf{Y}^* \mid \boldsymbol{\Omega}, \mathbf{U}^*, \mathbf{Q}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{V})$;

- Draw $\mathbf{Q}^*$ from $p(\mathbf{Q}^* \mid \mathbf{\Omega}, \mathbf{U}^*, \mathbf{Y}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{V})$;
- Draw $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta} \mid \mathbf{\Omega}, \mathbf{U}^*, \mathbf{Y}^*, \mathbf{Q}^*, \mathbf{U}, \mathbf{Z}, \mathbf{V})$.

Upon convergence, the posterior is approximated by the empirical distribution of the simulated observations. The convergence of the algorithm can be monitored by plotting the traces of estimates for different starting values or observing the values of EPSR [49] of unknown parameters. The technical details for implementing MCMC sampling are given in Appendix A.

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:

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

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

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

$$P(|\beta_j| < c_{\alpha/2}|\mathbf{U}, \mathbf{Z}, \mathbf{Y}) = 1 - \alpha, \; P(|\psi_j| < d_{\alpha/2}|\mathbf{U}, \mathbf{Z}, \mathbf{Y}) = 1 - \alpha \tag{28}$$

where α is any previously specified value in $(0,1)$. Calculation of the PCI can be achieved via the Monte Carlo method. For example, let $\{\beta_j^{(k)} : k = 1, \ldots, K\}$ 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 kth order statistics.

Another choice for variable determination in SS is based on the posterior probabilities of $f_{\beta j} = 1$ and $f_{\psi j} = 1$, which can be approximated by

$$\widehat{f}_{\beta j} = \frac{1}{K} \sum_{k=1}^{K} I\{f_{\beta j}^{(k)} = 1\}, \;\; \widehat{f}_{\psi j} = \frac{1}{K} \sum_{k=1}^{K} I\{f_{\psi j}^{(k)} = 1\}, \tag{29}$$

where $f_{\beta j}^{(k)}$ and $f_{\psi j}^{(k)} (k = 1, \ldots, K)$ are the k observations drawn from the posterior distribution via the Gibbs sampler. The variable w_j is selected in Parts One and Two if $\widehat{f}_{\beta j} > 0.5$ and $\widehat{f}_{\psi j} > 0.5$.

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 $y_{ij}(j = 1, \ldots, 6)$. We assume that s_i, ω_{ij} and y_{ij} satisfy Equations (1), (2), and (5) associated with (6), respectively, in which the number of fixed covariates is set to five. We generate x_{i1} and x_{i2} from the standard normal distribution, x_{i3} and x_{i4} from a binomial distribution with a probability of success of 0.3, and x_5 from a 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 the population parameters are set as follows: $\alpha = \gamma = 0.7$, $\boldsymbol{\beta} = (0.7, 0.0, 0.7, 0.0, 0.7, 0.0, 0.8)^T$, $\gamma = 0.7$, $\boldsymbol{\psi} = (0.7, 0.0, 0.7, 0.0, 0.7, 0.8, 0.0)^T$, $\sigma^2 = 1.0$,

$\boldsymbol{\mu} = 0.7 \times \mathbf{1}_6$, where $\mathbf{1}_6$ is a 6×1 vector for which the elements are in unity. The factor loading matrix $\boldsymbol{\Lambda}$ and covariance matrix $\boldsymbol{\Phi}$ are taken as

$$\boldsymbol{\Lambda}^T = \left[\begin{array}{cccccc} 1.0 & 0.8 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.8 & 0.8 \end{array} \right], \quad \boldsymbol{\Phi} = \left[\begin{array}{cc} 1.0 & 0.3 \\ 0.3 & 1.0 \end{array} \right], \tag{30}$$

in which the ones and zeros in $\boldsymbol{\Lambda}$ are treated as fixed to identify the model; the thresholds are set as $\delta_k = (-1.5^*, 0.0, 1.2, 2.5^*)^T$ for $k = 1, \ldots, 6$, 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 $N_2(\mathbf{0}, \boldsymbol{\Phi})$ and then drawing latent responses $\mathbf{Y}^*$ from (6). The the indicator responses $\mathbf{U}$, the intensity responses $\mathbf{Z}$ and the ordered categorical responses $\mathbf{Y}$, are sequentially generated from (1), (2) and (5). 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 the hyperparameters: for the parameters involved in the measurement model, we take $\boldsymbol{\mu}_0 = \mathbf{0}_6$ and $\boldsymbol{\Sigma}_0 = 100.0 \times \mathbf{I}_6$; the elements in $\boldsymbol{\Lambda}_0$ corresponding to the free parameters in $\boldsymbol{\Lambda}$ are set at zero, and $\mathbf{H}_{0k} = \mathbf{I}_2$ for $k = 1, \ldots, 6$; $\rho_0 = 10.0$, and $\mathbf{R}_0^{-1} = 6.0 \times \mathbf{I}_3$; for the threshold parameters $\boldsymbol{\delta}$, we take $\eta_1 = \cdots = \eta_5 = 1.0$, which denotes the uniform distribution of $\boldsymbol{p}$ on the simplex in $\mathbb{R}^5$; for the intercept parameters α, γ, and scale 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 $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are set as before. Note that these values can ensure that the corresponding priors are inflated enough; hence, we can expect this to enhance the robustness of the inference. In addition, we set $\nu_{\beta 0} = \nu_{\psi 0} = 0.001$ in Equations (11) and (12) 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 $\boldsymbol{\theta}$. Before the formal implementation, a few test runs are conducted as a pilot 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 values with sample size $n = 400$.
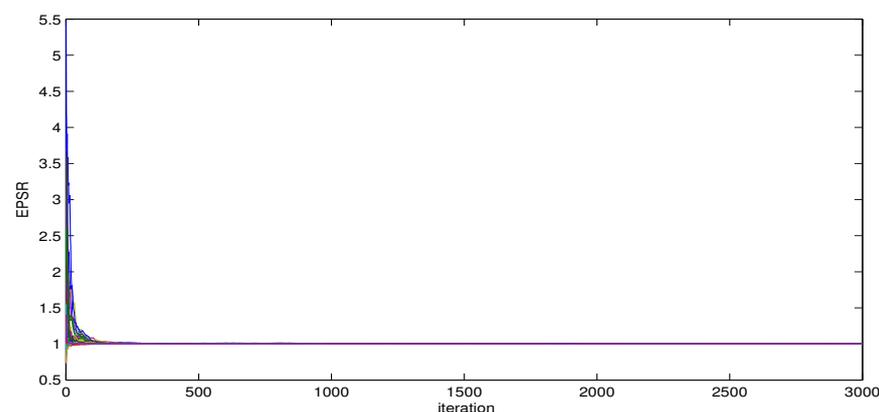
Figure 2. Plot of the values of EPSR of unknown parameters under three different starting values in which the colored solid lines represent the trajectories of EPSR of estimates against the number of iteraitons: simulation study and $n = 400$.

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 the burn-in phase and further collect 3000 observations for calculating the bias (BIAS), root

mean square (RMS), and standard deviation (SD) of the estimate across 100 replications. The BIAS and RMS of the j-th component $\widehat{\theta}_j$ in $\widehat{\boldsymbol{\theta}}$ are respectively defined as follows:

$$\text{BIAS}(\widehat{\theta}_j) = (\bar{\theta}_j - \theta_j^0), \bar{\theta}_j = \frac{1}{100}\sum_{\kappa=1}^{100}\widehat{\theta}_j^{(\kappa)}, \ \text{RMS}(\widehat{\theta}_j) = \sqrt{\frac{1}{100}\sum_{\kappa=1}^{100}(\widehat{\theta}_j^{(\kappa)} - \theta_j^0)^2}, \tag{31}$$

where θ_j^0 is the j-th element of the population parameters $\boldsymbol{\theta}^0$. The summaries of the estimates of the main parameters for the two scenarios are reported in Tables 1 and 2, where the sums of the SDs and RMSs 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	-0.020	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

Examination of Tables 1 and 2 gives the following findings: (i) Both methods produce satisfactory results. The performance of SS is slightly superior to that of BaLsso in terms of the sums of RMS and SD. 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) Except ψ_2 in Table 1 and ψ_4 in Table 2, for the regression coefficients and factor loadings with true values at zero, the RMSs and SDs produced by SS are uniformly smaller than those produced by BaLsso. This indicates that SS imposes more penalties on the regression coefficients and factor loadings than BaLsso to shrink them toward zero. (iii) For non-zero β_j and ψ_j, the situation becomes fiendishly complex. For example, the RMSs and SDs of the estimates of β_5, β_7, ψ_1, ψ_5, and ψ_6 in Table 1 show that the SS does a weaker job than BaLsso. This is also true for β_3 and ψ_3 in Table 2. The underlying reason perhaps lies in that the penalties imposed on the non-zero regression coefficients and factor loadings by SS are similar to those of BaLsso, and one is not overwhelmingly superior to the other. (iv) As expected, increasing the sample size improves the accuracy of the estimates for both SS and BaLsso.

Another simulation is conducted to assess the performance of the proposed method for 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 zero and covariance matrix $\mathbf{\Sigma}(7 \times 7)$ with $\Sigma_{jk} = \rho^{|j-k|}$, where Σ_{jk} is the $(j,k)^{th}$ entry of $\mathbf{\Sigma}$. We consider three scenarios for ρ: (i) $\rho = 0.1$, (ii) $\rho = 0.5$, and (iii) $\rho = 0.8$, which represent, respectively, weak, mild, and strong dependence among them. The values of $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are taken as $(1.0, 0.0, 1.0, 0.0, 1.0, 0.0, 1.0)^T$ and $(1.0, 0.0, 1.0, 0.0, 1.0, 1.0, 0.0)^T$, respectively, and the sample size is taken as $n = 1000$. The remainder of the model is set up the same as before. We implement MCMC sampling and collect 3000 observations after removing the first 2000 observations for posterior inference. We follow [43] and treat a regression coefficient as 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 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	1.290	-	1.247	1.335

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

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, the two methods exhibit satisfactory performances, with both having 100% correct rates across all situations; (ii) for zero regression coefficients, there are differences between the two methods, and SS uniformly outperforms BaLsso. The underlying reason perhaps is that for SS, the variances of estimates are set to be small enough to ensure that the coefficients are close to zero, while for BaLsso, the variances of estimates are controlled by the shrinkage parameters, which may not be large enough to ensure this point; (iii) with an increase in the strength of dependence, the correct rates of the two methods decrease. As an illustration, Figure 3 gives the plots of the correct rates of the selected variables for three scenarios; the values on the x-axis are the true values of β_1 to ψ_7 in the order given in Table 3.
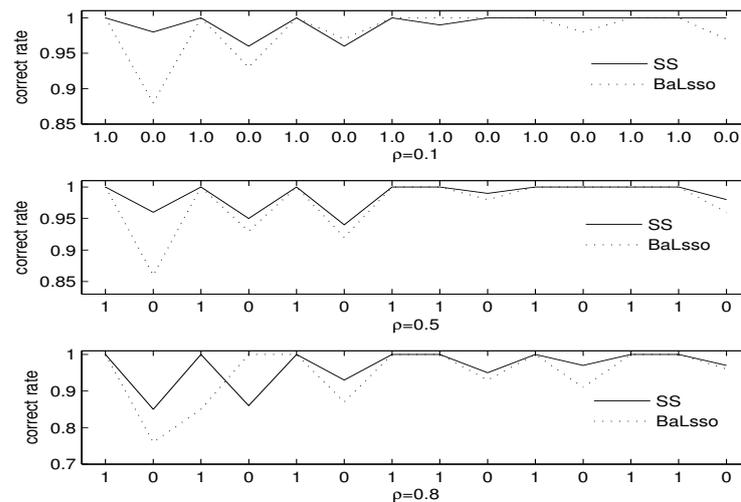
Figure 3. Plots of the correct rates of the selected variables under three scenarios: simulation study and $n = 1000$.

5. Chinese Household Financial 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 Chinese Household Financial Survey (CHFS), a non-profit institute organized by the Southeast University of Finance and Economics in Chengdu, Sichuan Province, China. The survey covers a series of questions that touch on information about various aspects of a household's financial situation. In this study, we only focus on the measurement 'gross debts per household (DEB)': the amount of secured debt and unsecured debt for the household under investigation. We extracted the data from a survey of Zhejiang Province in 2013. Due to some uncertain factors, some measurements of 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 DEB measurements 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 4 presents the histogram of DEB as well as the logarithms of positive values. It can be seen clearly that the dataset illustrates strong heterogeneity. The skewness and kurtosis of DEB are 1.1042 and 2.3361, respectively, which indicates that a single parametric model for DEB may be unappreciated.

We include the following measurements as the potential explanatory factors to interpret the variability of 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 these measurements. To unify the scale, all covariates were standardized.
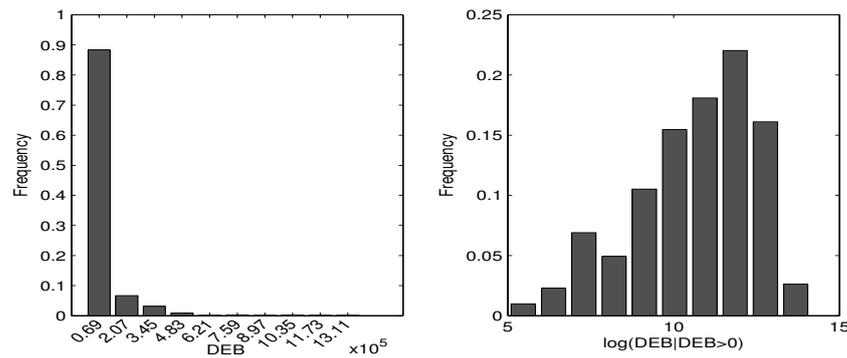
Figure 4. Histograms of DEB and the logarithms of their positive values: Chinese Household Financial Survey data. (**Left**) panel corresponds to DEB and (**right**) panel corresponds to $\log(\text{DEB}|\text{DEB} > 0)$.

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
Educational experience (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 of 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 as a latent factor into the current analysis. It is well-known that China is an ancient civilization with a long history, and Confucian culture is deeply rooted in social development. Economic activity or social development cannot be independent of cultural development. Hence, it is of practical interest to investigate how family culture affects behavior related to household finance debt. Based on the design of the questionnaire, we select the following three measurements as manifestations for η: (i) The preference for boys (BP, y_1): this is a three-category measurement coded as 0, 1, and 2, which correspond to the attitudes 'opposed', 'doesn't matter', and 'strongly support'; (ii) The attitude toward having a single child (SC), coded by 0, 1 and 2, according to the level of support. (iii) The importance of family in one's life: this measurement was originally based on a six-point scale (0 to 5) according to the support level. However, in view of the frequencies in the last three groups being small, we grouped them into three categories and recoded them as 0 (does not matter), 1 (important), and 2 (very important). In addition, as some manifestations are missing, we treat missing data as missing randomly and ignorable [50] and ignore the specific mechanics that result 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 collection 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 (6) and assume that η_i, *iid.* $\sim N(0, 1)$. The inputs of the 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 to be the same as those in the simulation study. To implement the MCMC sampling algorithm, we need to impute the missing data in $\mathbf{Y}$. This is done 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, that correspond to the missing entries $y_{ij,mis}$ 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 [51]

in the context of a latent variable model with polytomous data and fix δ_{j1} at $\Phi^{-1}(f_{j1}/n_j)$, where n_j is the size of $y_{obs,ij}$ that is 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 5). 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 their standard deviations.
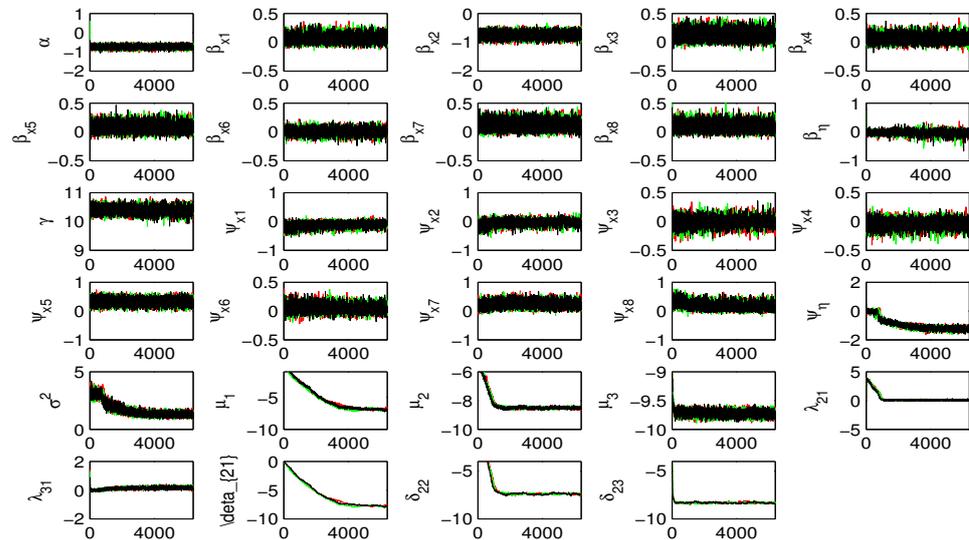
Figure 5. Trace plots of the estimates of unknown parameters against the number of iterations under SS prior, in which the colored solid lines repsent the traces of estimates under three starting values: CHFS data.

Table 5 gives the summary of two estimates of unknown parameters in the two-part model and factor analysis model. Examination of Table 5 shows that most of the estimates for both models are very close, but there exist 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, and 0.747 with standard deviations of 0.062, 0.070, and 0.072, respectively, while they equal 0.072, 0.082, and 0.092 with standard deviations of 0.07, 0.081, and 0.092 under BaLsso. These differences reflect the fact that the two methods impose different penalties on the regression coefficients during variable selection.

To see more clearly, Table 6 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', 'marital status', 'employment', 'number of adults', and 'family culture'. The two methods favor 'age','marital status', and 'number of adults' as helpful for improving model fit, while 'gender' and 'family culture' have less influence on the probability of holding household finance debt. However, there exist contradictory conclusions when selecting 'health condition', 'education', and 'income'. (ii) For Part Two, except for the factors 'age' and 'number of adults', the two methods give the same results. In particular, both methods support that 'family culture' is relevant to the amount of household finance debt being held. This fact is also revealed by [18] in an analysis of the CHFS using a two-part nonlinear latent variable model. Further interpretation is omitted in order to save space.

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

	SS		BaLsso			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					

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

	Part One		Part Two	
VAR	**SS**	**BaLsso**	**SS**	**BaLsso**
Gender	0	0	1	1
Age	1	1	1	0
Marital 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

A two-part latent variable model can be considered to be an extension of a traditional two-part model for situations where the latent variables are included to identify the unobserved heterogeneity of a population resulting from the absence of the observed covariates. When analyzing such a model, an important issue is to determine which factors are relevant to the outcome variable. This is especially true when the number of exogenous factors is high, because usual model selection/comparison procedures are extremely time-consuming. In this paper, we resort to a Bayesian variable selection method and develop a fully Bayesian variable selection procedure for semi-continuous data. Our formulation is along the lines of a spike and slab bimodal prior and recasts the distribution of regression coefficients and factor loadings as a hierarchy of priors over the parameter and model space. The selected variables are identified based on having a high posterior probability of occurrence. We also consider an adaptive Bayesian lasso (BaLsso) for reference. To facilitate computation, we recast the logistic regression model in Part One as a flavor of a normal mixture model by introducing latent Pólya–Gamma variables. This admits the conjugate full-conditional distributions for all regression coefficients, factor loadings, and factor variables.

Although Bayesian variable selection has its unique advantages, there are still some limitations that need to be considered with care. Firstly, 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 datasets. 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 do not conform to the model's convention, the performance of the method is poor. Therefore, these issues need to be carefully considered in real applications to ensure that the Bayesian SS method can be effectively applied to the specific dataset.

The proposed method can be applied to more general latent variable models, including multilevel SEMs [51] and longitudinal dynamic variable models with discrete variables [17,52]. These extensions are left for further study.

Author Contributions: Conceptualization, Q.Z.; methodology, Q.Z., Y.Z. and Y.X.; software, Q.Z. and Y.Z.; validation, Q.Z., Y.Z. and Y.X.; formal analysis, Q.Z., Y.Z. and Y.X.; investigation, Q.Z., Y.Z. and Y.X.; resources, Q.Z. and Y.X.; writing—original draft preparation, Y.Z. and Y.X.; writing—review and editing, Q.Z. and Y.X.; visualization, Y.Z.; supervision, Y.X.; project administration, Q.Z., Y.Z. and Y.X.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All the data included in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts 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 chain Monte Carlo
CHFS	Chinese Household Financial Survey

Appendix A

In this section, we present some technical details of the full conditionals for 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 $\mathbf{\Omega}$, $\mathbf{U}^*$, $\mathbf{Y}^*$, and $\boldsymbol{\theta}$ are exactly the same. The following derivations are mainly based on Bayes theorem.

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

It follows from Equations (2), (6) and (9), that

$$p(\mathbf{\Omega}|\cdots) = \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\cdots),$$

where

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

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

$$p(\boldsymbol{\omega}_i | \cdots) \overset{D}{=} N_m(\widehat{\boldsymbol{\mu}}_{\omega i}, \widehat{\boldsymbol{\Sigma}}_{\omega i}), \tag{A1}$$

where

$$\widehat{\boldsymbol{\mu}}_{\omega i} = \widehat{\boldsymbol{\Sigma}}_{\omega i} \left[\boldsymbol{\beta}_\omega \kappa_i^* + \boldsymbol{\psi}_\omega u_i z_i^* / \sigma^2 + \boldsymbol{\Lambda}^T (\mathbf{y}_i^* - \boldsymbol{\mu}) \right],$$

$$\widehat{\boldsymbol{\Sigma}}_{\omega i} = \left[\boldsymbol{\beta}_\omega \boldsymbol{\beta}_\omega^T u_i^* + \boldsymbol{\psi}_\omega \boldsymbol{\psi}_\omega^T u_i / \sigma^2 + \boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \boldsymbol{\Phi}^{-1} \right]^{-1}.$$

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

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

Following a similar derivation to that in [42], it can be shown that, given $\mathbf{U}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\theta}$, $\mathbf{U}^*$ is the Pólya–Gamma distribution through exponential tilting of the standard Pólya–Gamma density given by

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

where $\eta_i = \alpha + \boldsymbol{\beta}^T \mathbf{w}_i$. Drawing u_i from this distribution can be achieved via a rejection sampling method; see [42] or [53] for more details about this issue.

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

Note that

$$p(\mathbf{Y}^* | \cdots) \propto p(\mathbf{Y} | \mathbf{Y}^*, \boldsymbol{\delta}) p(\mathbf{Y}^* | \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$$

$$= \prod_{i=1}^n \prod_{k=1}^p \left(\sum_{\ell=0}^c I\{y_{ik} = \ell, \delta_{k\ell} < y_{ik}^* \leq \delta_{k\ell+1}\} \right) \times \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}(y_{ik}^* - \mu_k - \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i)^2 \right\}.$$

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

$$p(\mathbf{Y}^* | \cdots) = \prod_{i=1}^n \prod_{k=1}^p p(y_{ik}^* | \boldsymbol{\omega}_i, \boldsymbol{\theta}, y_{ik}),$$

$$p(y_{ik}^* | \boldsymbol{\omega}_i, \boldsymbol{\theta}, y_{ik}) = N(\mu_k + \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i, 1) I\{\delta_{k,y_{ik}} < y_{ik}^* \leq \delta_{k,y_{ik}+1}\}. \tag{A3}$$

This is the truncated normal distribution, and its drawing can be obtained via an inverse distribution sampling method; see, for example, [54].

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

Recall that $\boldsymbol{\theta}$ consists of α, $\boldsymbol{\beta}$, γ, $\boldsymbol{\psi}$, σ^2, $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\delta}$. Hence, the drawing of $\boldsymbol{\theta}$ can be accomplished by (i) drawing α from $p(\alpha | \cdots)$, (ii) drawing $\boldsymbol{\beta}$ from $p(\boldsymbol{\beta} | \cdots)$, (iii) drawing γ from $p(\gamma | \cdots)$, (iv) drawing $(\boldsymbol{\psi}, \sigma^2)$ from $p(\boldsymbol{\psi}, \sigma^2 | \cdots)$, (v) drawing $\boldsymbol{\mu}$ from $p(\boldsymbol{\mu} | \cdots)$, (vi) drawing $\boldsymbol{\Lambda}$ from $p(\alpha | \cdots)$, (vii) drawing $\boldsymbol{\Phi}$ from $p(\boldsymbol{\Phi} | \cdots)$, and (viii) drawing $\boldsymbol{\delta}$ from $p(\boldsymbol{\delta} | \cdots)$ sequentially. Note that given $\mathbf{U}^*$, $\mathbf{Y}^*$, and $\boldsymbol{\Omega}$, the models (2), (6) and (9), reduce to ordinary regression models and, hence, most of the full conditionals, similar to the regression coefficients and variance/covariance in the Bayesian regression analysis, are

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

$$p(\alpha|\cdots) = N(\widehat{\mu}_\beta, \widehat{\sigma}_\beta^2), \ \ p(\boldsymbol{\beta}|\cdots) = N_q(\widehat{\boldsymbol{\mu}}_\beta, \widehat{\boldsymbol{\Sigma}}_\beta), \tag{A4}$$

$$p(\gamma|\cdots) = N(\widehat{\mu}_\gamma, \widehat{\sigma}_\gamma^2), \ \ p(\boldsymbol{\psi}, \sigma^2|\cdots) = IG(\widehat{\alpha}_\sigma, \widehat{\beta}_\sigma) \times N_q(\widehat{\boldsymbol{\mu}}_\psi, \sigma^2 \widehat{\boldsymbol{\Sigma}}_\psi), \tag{A5}$$

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

$$p(\boldsymbol{\Phi}^{-1}|\cdots) = W_m(\rho + n, \widehat{\mathbf{R}}), \tag{A7}$$

in which

$$\widehat{\mu}_\alpha = \widehat{\sigma}_\alpha^2 \sum_{i=1}^n (\kappa_i - u_i^* \boldsymbol{\beta}^T \mathbf{w}_i), \ \widehat{\sigma}_\alpha^2 = (\sum_{i=1}^n u_i^* + \sigma_{\alpha 0}^{-2})^{-1},$$

$$\widehat{\boldsymbol{\mu}}_\beta = \widehat{\boldsymbol{\Sigma}}_\beta \sum_{i=1}^n \mathbf{w}_i (\kappa_i - \alpha u_i^*), \ \widehat{\boldsymbol{\Sigma}}_\beta^{-1} = \sum_{i=1}^n u_i^* \mathbf{w}_i \mathbf{w}_i + diag\{\gamma_\beta^{-2}\},$$

$$\widehat{\mu}_\gamma = \widehat{\sigma}_\gamma^2 \sum_{i=1}^n u_i (z_i - \boldsymbol{\psi}^T \mathbf{w}_i)/\sigma^2, \ \widehat{\sigma}_\gamma^2 = (\sum_{i=1}^n u_i/\sigma^2 + \sigma_{\gamma 0}^{-2})^{-1},$$

$$\widehat{\boldsymbol{\mu}}_\psi = \widehat{\boldsymbol{\Sigma}}_\psi \sum_{i=1}^n \mathbf{w}_i (z_i - \gamma) u_i/\sigma^2, \ \widehat{\boldsymbol{\Sigma}}_\psi^{-1} = \sum_{i=1}^n u_i \mathbf{w}_i \mathbf{w}_i + diag\{\gamma_\psi^{-2}\},$$

$$\widehat{\alpha}_\sigma = a_0 + |\mathcal{I}|/2,$$

$$\widehat{\beta}_\sigma = b_0 + \frac{1}{2}(\sum_{i=1}^n u_i z_i^2 - \widehat{\boldsymbol{\mu}}_\psi^T \widehat{\boldsymbol{\Sigma}}_\psi^{-1} \widehat{\boldsymbol{\mu}}_\psi + \boldsymbol{\Lambda}_{0k}^T \mathbf{H}_{0k}^{-1} \boldsymbol{\Lambda}_{0k}),$$

$$\widehat{\mathbf{m}}_\mu = \widehat{\boldsymbol{\Sigma}}_\mu (\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + n(\bar{\mathbf{Y}}^* - \boldsymbol{\Lambda}\bar{\boldsymbol{\Omega}})), \ \widehat{\boldsymbol{\Sigma}}_\mu^{-1} = n\mathbf{I}_p + \boldsymbol{\Sigma}_0^{-1},$$

$$\widehat{\boldsymbol{\Lambda}}_k = \widehat{\mathbf{H}}_k (\mathbf{H}_{0k}^{-1} \boldsymbol{\Lambda}_{0k} + \boldsymbol{\Omega}^T \mathbf{Y}_{[k]}^{**}), \ \widehat{\mathbf{H}}_k^{-1} = n\boldsymbol{\Phi}^{-1} + \boldsymbol{\Omega}^T \boldsymbol{\Omega},$$

$$\widehat{\mathbf{R}}^{-1} = \mathbf{R}_0^{-1} + \boldsymbol{\Omega}^T \boldsymbol{\Omega},$$

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

However, for δ, we note that

$$p(\boldsymbol{\delta}|\cdots) = \prod_{k=1}^p p(\boldsymbol{\delta}_k|\mathbf{Y}_{[k]}^*, \mathbf{Y}_{[k]}), \ \text{and}$$

$$p(\boldsymbol{\delta}_k|\mathbf{Y}_{[k]}^*, \mathbf{Y}_{[k]}) \propto p(\boldsymbol{\delta}_k) \prod_{i=1}^n \prod_{\ell=0}^c I\{y_{ik} = \ell, \delta_{k\ell} < y_{ik}^* \le \delta_{k,\ell+1}\}.$$

Hence, drawing of δ can be obtained by drawing $\boldsymbol{\delta}_k$ from $p(\boldsymbol{\delta}_k|\cdots)$ independently. Moreover, under prior (25), it can be shown that

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

where $\boldsymbol{\delta}_{k,(-\ell)}$ is the vector of $\boldsymbol{\delta}_k$ 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 (26) that

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

where

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

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 δ_{kl} via an 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 drawing of the truncated beta distribution can be obtained by implementing an inverse distribution sampling method.

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

First of all, it is noted that $\mathbf{Q}^*$ consists of $\mathbf{F}_\beta$, $\mathbf{F}_\psi$, w_β, w_ψ, η_β^{-2}, and η_ψ^{-2} under SS, and it is 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

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

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

which indicates that the components in the posteriors of $\mathbf{F}_\beta$ and $\mathbf{F}_\psi$ are independent. Further, it follows easily from (13) that

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

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

where

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

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

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

Secondly, it is noted that

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

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

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

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

Hence,

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

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

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

Lastly, it follows from

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

that

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

where

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

For BaLasso, we follow the practice in [38] and can show

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

in which

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

where $IG(\mu, \lambda)$ is an inverse Gaussian distribution with density $\sqrt{\lambda/(2\pi)}x^{-3/2}\exp\{-\lambda(x-\mu)^2/(2\mu^2 x)\}(x > 0)$ [55].

Similarly,

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

in which

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

References

1. Deb, P.; Munkin, M.K.; Trivedic, R.K. Bayesian analysis of the two-part model with endogeneity: Application to health care expenditure. *J. Appl. Econ.* **2006**, *21*, 1081–1099. [CrossRef]

2. Cragg, J.G. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* **1971**, *39*, 829–844. [CrossRef]
3. Neelon, B.; Zhu, L.; Neelon, S.E.B. Bayesian two-part spatial models for semicontinuous data with application to emergency department expenditures. *Biostatistics* **2015**, *16*, 465–479. [CrossRef]
4. Manning, W.G.; Morris, C.N.; Newhouse, J.P.; Orr, L.L.; Duan, N.; Keeler, E.B.; Leibowitz, A. A two-part model of the demand for medical Care: Preliminary results from the health insurance experiment. In *Health, Economics, and Health Economics*; van der Gaag, J., Perlman, M., Eds.; North-Holland: Amsterdam, The Netherlands, 1991; pp. 103–104.
5. Su, L.; Tom, B.D.; Farewell, V.T. Bias in 2-part mixed models for longitudinal semi-continuous data. *Biostatistics* **2009**, *10*, 374–389. [CrossRef] [PubMed]
6. Su, L.; Tom, B.D.; Farewell, V.T. A likelihood-based two-part marginal model for longitudinal semi-continuous data. *Stat. Methods Med. Res.* **2015**, *24*, 194–205. [CrossRef] [PubMed]
7. Duan, N.; Manning, W.G.; Morris, C.N.; Newhouse, J.P. A comparison of alternative models for the demand for medical Care. *J. Bus. Econ. Stat.* **1983**, *1*, 115–126.
8. Liu, L.; Cowen, M.E.; Strawderman, R.L.; Shih, Y.C.T. A flexible two-part random effects model for correlated medical costs. *J. Health Econ.* **2010**, *29*, 110–123. [CrossRef] [PubMed]
9. Smith, V.A.; Neelon, B.; Preisser, J.S.; Maciejewski, L. A marginalized two-part model for semicontinuous data. *Stat. Med.* **2015**, *33*, 4891–4903. [CrossRef] [PubMed]
10. Tooze, J.A.; Grunwald, J.K.; Jones, R.H. Analysis of repeated measures data with clumping at zero. *Stat. Methods Med. Res.* **2002**, *11*, 341–355. [CrossRef] [PubMed]
11. Brown, R.A.; Monti, P.M.; Myers, M.G.; Martin, R.A.; Rivinus, T.; Dubreuil, M.E.T.; Rohsenow, D.J. Depression among cocaine abusers in treatment: Relation to cocaine and alcohol use and treatment outcome. *Am. J. Psychiat.* **1998**, *155*, 220–225. [CrossRef]
12. Olsen, M.K.; Schafer, J.L. A two-part random-effects model for semicontinuous longitudinal data. *J. Am. Stat. Assoc.* **2001**, *96*, 730–745. [CrossRef]
13. Xing, D.Y.; Huang, Y.X.; Chen, H.N.; Zhu, Y.L.; Dagen, G.A.; Baldwin, J. Bayesian inference for two-part mixed effects model using skew distributions, with application to longitudinal semi-continuous alcohol data. *Stat. Methods Med. Res.* **2017**, *26*, 1838–1853. [CrossRef]
14. Chen, J.Y.; Zheng, L.Y.; Xia, Y.M. Bayesian analysis for two-part latent variable model with application to fractional data. *Commun. Stat. Theory Methods* 2023, *preprint*.
15. Kim, Y.; Muthén, B.O. Two-part factor mixture modeling: Application to an aggressive behavior measurement instrument. *Struct. Equ. Model. Multidiscip. J.* **2009**, *16*, 602–624. [CrossRef] [PubMed]
16. Feng, X.; Lu, B.; Song, X.Y.; Ma, S. Financial literacy and household finances: A Bayesian two-part latent variable modeling approach. *J. Empir. Financ.* **2019**, *51*, 119–137. [CrossRef]
17. Xia, Y.M.; Tang, N.S. Bayesian analysis for mixture of latent variable hidden Markov models with multivariate longitudinal data. *Comput. Stat. Data Anal.* **2019**, *132*, 190–211. [CrossRef]
18. Gou, J.W.; Xia, Y.M.; Jiang, D.P. Bayesian analysis of two-part nonlinear latent variable model: Semiparametric method. *Stat. Model* **2023**, *23*, 721–741. [CrossRef]
19. Xiong, S.C.; Xia, Y.M.; Lu, B. Bayesian analysis of two-part latent variable model with mixed data. *Commun. Math. Stat.* 2023, *preprint*. [CrossRef]
20. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]
21. Fu, W.J. Penalized regression: The bridge versus the lasso. *J. Comput. Graph. Stat.* **1998**, *7*, 109–148.
22. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2009.
23. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity—The Lasso and Generalization*; CRC Press: New York, NY, USA, 2015.
24. Kuo, L.; Mallick, B.K. Variable selection for regression models. *Sankhyā Indian J. Stat. Ser. B* **1998**, *60*, 65–81.
25. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [CrossRef]
26. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
27. Zou, H. The adaptive Lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [CrossRef]
28. Zhang, W.; Ota, T.; Shridhar, V.; Chien, J.; Wu, B.; Kuang, R. Network based survival analysis reveals subnetwork signatures for predicting outcomes of ovarian cancer treatment. *PLoS Comput. Biol.* **2013**, *9*, e1002975. [CrossRef]
29. Zhao, Q.; Shi, X.J.; Xie, Y.; Huang, J.; Shia, B.C.; Ma, S. Combining multidimensional genomic measurements for predicting cancer prognosis: Observations from TCGA. *Brief. Bioinform.* **2015**, *16*, 291–303. [CrossRef]
30. George, E.I.; McCulloch, R.E. Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.* **1993**, *88*, 881–889. [CrossRef]
31. George, E.I.; McCulloch, R.E. Approaches for Bayesian variable selection. *Stat. Sin.* **1997**, *7*, 339–373.
32. Chipman, H.A. Bayesian variable selection with related predictors. *Canad. J. Statist.* **1996**, *24*, 17–36. [CrossRef]
33. Ishwaran, H.; Rao, J.S. Spike and Slab gene selcetion for multigroup microarray data. *J. Am. Stat. Assoc.* **2005**, *87*, 371–390.
34. Ishwaran, H.; Rao, J.S. Spike and Slab variable selection: Frequentist and Bayesian strageies. *Ann. Stat.* **2005**, *33*, 730–773. [CrossRef]

35. Mitchell, T.J.; Beauchamp, J.J. Bayesian variable selection in linear regression. *J. Am. Stat. Assoc.* **1988**, *83*, 1023–1032. [CrossRef]

36. Rockova, V.; George, E.I. EMVS: The EM approach to Bayesian variable selection. *J. Am. Stat. Assoc.* **2014**, *109*, 828–846. [CrossRef]

37. Tang, Z.X.; Shen, Y.P.; Zhang, X.Y.; Yi, N.J. The Spike-and-Slab Lasso generalized linear modelsfor prediction and associated genes detection. *Genetics* **2017**, *205*, 77–88. [CrossRef]

38. Park, T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [CrossRef]

39. Skrondal, A.; Rabe-Hesketh, S. *Generalized Latent Variable Modelling: Multilevel, Longitudinal and Structural Equation Models*; Chapman & Hall/CRC: London, UK, 2004.

40. Bollen, K.A. *Structural Equations with Latent Variables*; John Wiley & Sons: New York, NY, USA, 1989.

41. Lee, S.Y. *Structural Equation Modeling: A Bayesian Approach*; John Wiley & Sons: New York, NY, USA, 2007.

42. Polson, N.G.; Scott, J.G.; Windle, J. Bayesian inference for logistic models using Polya-Gamma latent variables. *J. Am. Stat. Assoc.* **2013**, *108*, 1339–1349. [CrossRef]

43. Feng, X.; Wang, Y.F.; Lu, B.; Song, X.Y. Bayesian regularized quantile structural equation models. *J. Multivar. Anal.* **2017**, *154*, 234–248. [CrossRef]

44. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*; John Wiley & Sons: New York, NY, USA, 1984.

45. Sha, N.J.; Dechi, B.O. A Bayes inference for ordinal response with latent variable approach. *Stats* **2019**, *2*, 321–331. [CrossRef]

46. Tanner, M.A.; Wong, W.H. The calculation of posterior distributions by data augmentation (with discussion). *J. Am. Stat. Assoc.* **1987**, *82*, 528–550. [CrossRef]

47. Gelfand, A.E.; Smith, A.F.M. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **1990**, *85*, 398–409. [CrossRef]

48. Geman, S.; Geman, D. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *PAMI-6*, 721–741. [CrossRef]

49. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* **1992**, *7*, 457–511. [CrossRef]

50. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons: New Yor, NY, USA, 2002.

51. Song, X.Y.; Lee, S.Y. A tutorial on the Bayesian approach for analyzing structural equation models. *J. Math. Psychol.* **2012**, *56*, 135–148. [CrossRef]

52. Song, X.Y.; Xia, Y.M.; Zhu, H.T. Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics* **2017**, *73*, 313–323. [CrossRef] [PubMed]

53. Devroye, L. *Non-Uniform Random Variate Generation*; Springer: New York, NY, USA, 1986.

54. Ross, S.M. *A Course in Simulation*; MacMillan: New York, NY, USA, 1991.

55. Chhikara, R.S.; Folks, L. *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*; Marcel Dekker: New York, NY, USA, 1989.