

## Research Article

# Assessing Heterogeneity for Factor Analysis Model with Continuous and Ordinal Outcomes

**Ye-Mao Xia and Jian-Wei Gou**

*Department of Applied Mathematics, Nanjing Forestry University, Nanjing, Jiangsu 210037, China*

Correspondence should be addressed to Ye-Mao Xia; [ym\\_xia71@163.com](mailto:ym_xia71@163.com)

Received 8 December 2015; Revised 23 February 2016; Accepted 2 March 2016

Academic Editor: Wei-Chiang Hong

Copyright © 2016 Y.-M. Xia and J.-W. Gou. 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.

Factor analysis models with continuous and ordinal responses are a useful tool for assessing relations between the latent variables and mixed observed responses. These models have been successfully applied to many different fields, including behavioral, educational, and social-psychological sciences. However, within the Bayesian analysis framework, most developments are constrained within parametric families, of which the particular distributions are specified for the parameters of interest. This leads to difficulty in dealing with outliers and/or distribution deviations. In this paper, we propose a Bayesian semiparametric modeling for factor analysis model with continuous and ordinal variables. A truncated stick-breaking prior is used to model the distributions of the intercept and/or covariance structural parameters. Bayesian posterior analysis is carried out through the simulation-based method. Blocked Gibbs sampler is implemented to draw observations from the complicated posterior. For model selection, the logarithm of pseudomarginal likelihood is developed to compare the competing models. Empirical results are presented to illustrate the application of the methodology.

## 1. Introduction

Owing to its wide applications in behavioral and social science researches, analysis of factor analysis models with mixed data structure has received a lot of attention; see [1–6]. However, most of these methods are mainly developed within particular parametric distribution families such as the exponential family or normal scale mixture family, which have a limited role in dealing with the distributional deviations, in particular heterogeneity or multimodality of the data. Though some robust methods are developed to downweight the influence of the outliers [7–12], most of them are still confined to dealing with unimodality and are less effective for the asymmetric and/or multimodal problems.

Recently, some authors focused on the Bayesian semiparametric modeling for latent variables model. For multivariate categorical data analysis, Kottas et al. [13] extended the traditional multivariate probit model [14–16] to a flexible underlying prior probability model. The usual single multivariate normal model for the latent variables is replaced by a mixture of normal priors with infinite number of

components. And, for the latent variable model with fixed covariates and continuous responses, Lee et al. [17] established the semiparametric Bayesian hierarchical model for the structural equation models (SEMs) by relaxing the common normal distribution of exogenous factors to follow a finite-dimensional Dirichlet process [18]. Song et al. [19] developed a semiparametric Bayesian procedure for analyzing the latent variable model with unordered categorical data. For some recent advances in semiparametric analysis for factor analysis model, see [20–23] among others.

In this paper, we developed a Bayesian semiparametric approach for analyzing factor analysis model with mixed continuous and ordinal responses. The methods are twofold. Firstly, we extended Kottas, Müller, and Quintana's model to a more general multivariate model which contains factor variables. This extension aims to interpret the relationships between measurements and latent variables and explore correlations among the multiple manifest variables. Moreover, we treat the threshold parameters as unknown and estimate them simultaneously with other model parameters, thus providing a more flexible approach to fit the data. Secondly,

we introduce the truncated Dirichlet process prior as the prior of the mean vector and variance-covariance parameters of unique errors and latent variables. This facilitates the interpretation of heterogeneity in the mean and/or covariance structure across the subjects.

This paper is organized as follows. We first introduce the Bayesian semiparametric modeling framework for factor analysis model with continuous and ordinal variables. We then present the Markov chain Monte Carlo procedure for parameters estimation and model selection. Simulation studies and a real example are provided to illustrate the performance of the proposed procedure. We close with some remarks and concluding comments.

## 2. Model Description

**2.1. Factor Analysis Model with Continuous and Ordinal Responses.** Suppose that a  $p$ -dimensional mixed observed vector  $y_i = (x_i^T, z_i^T)^T$  contains  $r$  continuous variables  $x_i = (x_{i1}, \dots, x_{ir})^T$  and  $s = p - r$  ordinal variables  $z_i = (z_{i1}, \dots, z_{is})^T$  with  $z_{ij}$  taking an integral value in  $\mathbb{S}_j = \{0, 1, \dots, b_j\}$  for  $j = 1, \dots, s, i = 1, \dots, n$ . We assume that the observed ordinal vector  $z_i$  is related to the unobserved continuous vector  $u_i = (u_{i1}, \dots, u_{is})^T$  through

$$z_{ij} = l \quad \text{if } \tau_{jl} < u_{ij} \leq \tau_{j,l+1}, \quad l \in \mathbb{S}_j, \quad (1)$$

where  $\{\tau_{jl} : l = 0, \dots, b_j, j = 1, \dots, s\}$  is a set of unknown threshold parameters that define the categories:  $-\infty = \tau_{j0} < \tau_{j1} < \dots < \tau_{jb_j} < \tau_{j,b_j+1} = \infty$ . Hence, for the  $j$ th variable  $z_{ij}$ , there are  $b_j + 1$  categories.

Let  $y_i^* = (x_i^T, u_i^T)^T$  denote the vector of continuous observed measurements and unobserved variables. For subject  $i$ , we formulate the dependence among  $y_i^*$ 's through the following measurement model:

$$y_i^* = \mu + \Lambda \omega_i + \epsilon_i, \quad (2)$$

where  $\mu$  is a  $p \times 1$  intercept vector,  $\Lambda$  is a  $p \times m$  factor loading matrix,  $\omega_i$  is an  $m \times 1$  vector of latent variables, and  $\epsilon_i$  is a  $p \times 1$  vector of measurement errors which is independent of  $\omega_i$ . In many applications,  $\omega_i$  may represent the hypothesized factors underlying manifest responses and/or unobserved heterogeneity not explained by covariates.

The latent variable model with mixed continuous and ordinal responses defined by (1) and (2) faces two sources of identification problems. The first one is associated with the determinacy of latent variables  $y^*$  in modeling of categorical variables, and the second one is related to the uniqueness of the factor loadings matrix. To solve the first problem, we use the common method [24] to fix endpoints  $\tau_{j1}$  and  $\tau_{jb_j}$  ( $j = 1, \dots, s$ ) at preassigned values. For the second problem, we follow the usual practice in structural equation modeling to identify the covariance matrix of  $y_i^*$  by fixing appropriate elements in  $\Lambda$  at preassigned values.

Let  $\theta$  be the parametric vector formed by the unknown parameters contained in  $\{\mu, \Psi_e, \Phi\}$  and let  $\vartheta$  denote the free parameters contained in factor loading matrices  $\Lambda$  and

$\tau = (\tau_1^T, \dots, \tau_s^T)^T$  with  $\tau_j = (\tau_{j1}, \dots, \tau_{jb_j})^T$ . Based on the assumptions of (2), the conditional distribution of  $y_i^*$  given  $(\theta, \vartheta)$  is a normal distribution with mean vector  $\mu$  and covariance matrix  $\Sigma(\theta, \vartheta) = \Lambda \Phi \Lambda^T + \Psi_e$ .

Note that the latent factors here play an important role in characterizing the associations between the observed variables. It can be seen clearly that  $z_i$  and  $x_i$  are dependent when  $\omega_i$  is integrated out. The marginal density of  $y_i$  is given by

$$\begin{aligned} p(y_i | \vartheta, \theta) \\ = \int p(x_i | \omega_i, \vartheta, \theta) p(z_i | \omega_i, \vartheta, \theta) p(\omega_i | \vartheta, \theta) d\omega_i \end{aligned} \quad (3)$$

with

$$\begin{aligned} p(z_i | \omega_i, \vartheta, \theta) = \prod_{j=1}^s \left[ \Phi_c \left( \frac{\tau_{jz_{ij}+1} - \mu_{r+j} - \Lambda_{r+j}^T \omega_i}{\Psi_{\epsilon r+j}} \right) \right. \\ \left. - \Phi_c \left( \frac{\tau_{jz_{ij}} - \mu_{r+j} - \Lambda_{r+j}^T \omega_i}{\Psi_{\epsilon r+j}} \right) \right], \end{aligned} \quad (4)$$

in which  $\Phi_c(\cdot)$  is the standard normal cumulative distribution function.

**2.2. Bayesian Semiparametric Hierarchical Modeling.** Let  $p(y_i^* | \theta, \vartheta)$  be the conditional density of  $y_i^*$  given  $(\theta, \vartheta)$  and denote by  $F$  a prior distribution function of  $\theta$ . Suppose that  $F$  is proper; we define the following mixture density:

$$p(y_i^* | F, \vartheta) = \int p(y_i^* | \theta, \vartheta) F(d\theta), \quad (5)$$

in which  $F(d\theta)$  is the conditional distribution of  $\theta$  given  $F$ . By taking a prior for  $\vartheta$  and restricting  $F$  to be a parametric family of distributions indexed by  $\theta$ , we complete the Bayesian parametric model specification. However, this restriction severely constrains the estimation of  $\theta$  and produces estimators that shrink data values toward the same points. A more flexible modeling for  $y_i^*$  is to treat  $F$  as random and assign a prior for it. For this end, we introduce a latent variable vector  $\theta_i = \{\mu_i, \Psi_{\epsilon i}, \Phi_i\}$  and assume that, given  $\theta_i$ ,  $y_i^*$ 's are conditionally independent and drawn from  $p(y_i^* | \theta_i, \vartheta)$ . Furthermore, we suppose that  $\theta_i$ 's are independent and identically distributed (i.i.d.) according to  $F$  with a prior  $\mathcal{P}$  on it. As a result, we break the mixture model  $p(y_i^* | F, \vartheta)$  into

$$\begin{aligned} [y_i^* | \theta_i, \vartheta] &\stackrel{\text{iid}}{\sim} p(y_i^* | \theta_i, \vartheta), \\ [\theta_1, \dots, \theta_n | F] &\stackrel{\text{iid}}{\sim} F, \quad F \sim \mathcal{P}, \end{aligned} \quad (6)$$

where ‘‘ind’’ means ‘‘independent’’ and  $\mathcal{P}$  is a prior of  $F$ .

We consider the following truncated version of Dirichlet process for  $F$ :

$$\mathcal{P}(\cdot) = \mathcal{P}_G(\cdot) = \sum_{k=1}^G \pi_k \delta_{\theta_k^*}(\cdot), \quad (7)$$

in which  $\delta_{\theta_k^*}(\cdot)$  denotes a discrete probability measure concentrated on atom  $\theta_k^*$  and  $\pi_k$  ( $k = 1, \dots, G$ ), independent of  $\theta_k^*$ , are random weights constructed through the following stick-breaking procedure:

$$\begin{aligned} \pi_1 &= V_1, \\ \pi_k &= (1 - V_1) \cdots (1 - V_{k-1}) V_k, \quad k = 2, \dots, G - 1, \\ \pi_G &= (1 - V_1) \cdots (1 - V_{G-1}), \end{aligned} \quad (8)$$

with  $V_k \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(1, \alpha)$ ;  $\theta_k^*$ 's are i.i.d with common distribution  $F_0$ .

Truncated Dirichlet process prior (7) can be considered as a truncation version of Dirichlet process [25–30] in the nonparametric Bayesian analysis. It can be shown that, under (7) and (8), for any Borel set  $A$  in  $\mathbb{R}^p$ ,

$$\begin{aligned} \mathbb{E}F(A) &= F_0(A), \\ \text{Var}(F(A)) &= \frac{F_0(A)(1 - F_0(A))}{\alpha + 1} \left(1 + \frac{\alpha}{G}\right). \end{aligned} \quad (9)$$

This indicates that  $F_0$  can be served as the starting point or guess of  $F$  and  $\alpha$  determines the concentration of the prior around  $F_0$ . In practice, the value of  $G$  is either set to a large, predetermined value (e.g.,  $G \geq 100$ ) or chosen empirically. For instance, Ishwaran and Zarepour [31] suggested that the adequacy of the truncation level,  $G$ , can be assessed by evaluating moments of the tail probability. Our simulation results have shown that  $G = 100$  is more than adequate for the model considered in the present context.

Now, we specify the distribution  $F_0$ . Recalling that by convention  $\theta_k^*$  is the collection of  $\{\mu_k^*, \Psi_{\epsilon k}^*, \Phi_k^*\}$ , hence, we assume that

$$\begin{aligned} &F_0(\mu_k^*, \Psi_{\epsilon k}^*, \Phi_k^* \mid \nu, \Sigma_\nu, R) \\ &= N(\mu_k^* \mid \nu, \Sigma_\nu) \cdot \prod_{j=1}^m \text{Gamma}^{-1}(\Psi_{\epsilon k j}^* \mid \alpha_{\epsilon 0 j}, \beta_{\epsilon 0 j}) \\ &\quad \cdot \text{Wishart}^{-1}(\Phi_k^* \mid \rho_0, R^{-1}), \end{aligned} \quad (10)$$

where  $\nu, \Sigma_\nu$ , and  $R$  are hyperparameters,  $\Sigma_\nu = \text{diag}\{\sigma_{\nu 1}, \dots, \sigma_{\nu p}\}$  is a diagonal matrix with the  $k$ th diagonal element  $\sigma_{\nu k}$ , and  $R$  is an  $m_2 \times m_2$  positive definite matrix;  $\text{Gamma}^{-1}(\alpha_{\epsilon 0 j}, \beta_{\epsilon 0 j})$  refers to the inverse gamma distribution with shaper parameters  $\alpha_{\epsilon 0 j}$  and scale parameters  $\beta_{\epsilon 0 j}$ , respectively, and  $\text{Wishart}^{-1}$  denotes the inverse Wishart distribution [32].

Modeling  $F$  in (7) into the random probability measure and incorporating the latent variable  $\omega_i$  into (5) generate the following hierarchical model: for  $i = 1, \dots, n$ ,

$$\begin{aligned} (y_i^* \mid \omega_i, \theta_i, \vartheta) &\stackrel{\text{i.i.d.}}{\sim} N(\mu_i + \Lambda \omega_i, \Psi_{\epsilon i}), \\ (\omega_i \mid \theta_i) &\stackrel{\text{i.i.d.}}{\sim} N(0, \Phi_i), \\ (\theta_i \mid F) &\stackrel{\text{i.i.d.}}{\sim} F, \quad F \sim \mathcal{P}_G(\cdot), \end{aligned} \quad (11)$$

where  $\mathcal{P}_G$  is given by (7) and (8).

### 3. Parameters Estimation and Model Selection

*3.1. Prior Specifications and Estimation via Blocked Gibbs Sampler.* Let  $\Theta^* = \{\theta_k^* : k = 1, \dots, G\}$ . To implement Bayesian analysis, blocked Gibbs sampler is used to simulate observations from the posterior. The key for blocked Gibbs sampler is to recast model (11) completely by introducing the cluster variables  $L = (L_1, \dots, L_n)^T$  such that  $\theta_i = \theta_{L_i}^*$ . Consequently, the semiparametric hierarchical model (11) can be reformulated as the following framework:

$$\begin{aligned} (y_i^* \mid \omega_i, \theta_i, \vartheta) &\stackrel{\text{i.i.d.}}{\sim} N(\mu_i + \Lambda \omega_i, \Psi_{\epsilon i}), \\ (\omega_i \mid \theta_i) &\stackrel{\text{i.i.d.}}{\sim} N(0, \Phi_i), \\ (L_i = \cdot \mid \pi) &\stackrel{\text{i.i.d.}}{\sim} \sum_{k=1}^G \pi_k \delta_k(\cdot), \\ (\pi, \Theta^*) &\sim p(\pi) p(\Theta^*), \\ \vartheta &\sim p(\vartheta), \\ \tau &\sim p(\tau), \end{aligned} \quad (12)$$

where  $p(\vartheta)$  is a prior of  $\vartheta$ ,  $p(\pi)$  is the stick-breaking prior given by (8) with  $[V_i \mid \alpha] \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(1, \alpha)$ , and  $p(\Theta^*)$  is the joint distribution of  $\Theta^*$  given by

$$\begin{aligned} p(\Theta^* \mid \nu, \Sigma_\nu, R) &= \prod_{k=1}^G p(\theta_k^* \mid \nu, \Sigma_\nu, R) \\ &= \prod_{k=1}^G p(\mu_k^*, \Psi_{\epsilon k}^*, \Phi_k^* \mid \nu, \Sigma_\nu, R), \\ &[\mu_k^*, \Psi_{\epsilon k}^*, \Phi_k^* \mid \nu, \Sigma_\nu, R] \stackrel{\text{i.i.d.}}{\sim} F_0 \end{aligned} \quad (13)$$

in which  $F_0(\cdot \mid \nu, \Sigma_\nu, R)$  is given in (10).

For the Bayesian analysis, we need to specify priors for the parameters involved in the model. The whole parameters can be divided into two parts: parametric component part  $\{\vartheta, \tau\}$  and nonparametric component part  $\{\nu, \Sigma_\nu, R, \alpha\}$ . For the parametric components, we assume that  $p(\vartheta, \tau) = p(\vartheta)p(\tau)$  with

$$\begin{aligned} p(\Lambda_k) &\stackrel{D}{=} N(\Lambda_{0k}, H_{\epsilon 0k}), \\ p(\tau) &= \prod_{j=1}^s p(\tau_j) = \prod_{j=1}^s p(\tau_{j,2}, \dots, \tau_{j,b_j-1}) \\ &\propto \prod_{j=1}^s I\{\tau_{j,2} < \dots < \tau_{j,b_j-1}\}, \end{aligned} \quad (14)$$

where  $\Lambda_k$  is a  $p \times 1$  column vector that contains unknown parameters in the  $k$ th row of  $\Lambda$ .

For the hyperparameter  $\beta = \{\nu, \Sigma, R, \alpha\}$ , we consider the following conjugate priors:

$$\begin{aligned} \nu &\sim N(\mu_0, \Sigma_0), \\ R &\sim \text{Wishart}^{-1}(\rho_0^\phi, R_0^\phi), \\ \sigma_{\nu k} &\sim \text{Gamma}^{-1}(\kappa_1, \kappa_2), \\ \alpha &\sim \text{Gamma}(\tau_1, \tau_2). \end{aligned} \quad (15)$$

The hyperparameters  $\mu_0, \Sigma_0, \Lambda_{0k}, H_{\epsilon 0k}, R_0^\phi, \alpha_{\epsilon 0j}, \beta_{\epsilon 0j}, \rho_0, \rho_0^\phi, \kappa_1, \kappa_2, \tau_1$ , and  $\tau_2$  in (10), (14), and (15) are treated as known.

Let  $Y(n \times p) = (y_1, \dots, y_n)^T$ ,  $\Omega = (\omega_1, \dots, \omega_n)^T$ , and  $Y^* = (y_1^*, \dots, y_n^*)^T$ . Posterior analysis in relation to the complex  $p(\vartheta, \beta \mid Y)$  is carried out through the data augmentation technique [33]. Specifically, we treat the latent quantities  $\{\Omega, Y^*, \pi, \Theta^*, L\}$  as missing data and augment them with the observed data. A sequence of random observations is generated from the joint posterior distribution  $p(\Omega, Y^*, \vartheta, \pi, \Theta^*, L, \beta \mid Y)$  by the blocked Gibbs sampler [31, 34], coupled with the Metropolis-Hastings algorithm [35, 36]: given  $\{\Omega^{(l)}, Y^{*(l)}, \pi^{(l)}, \Theta^{*(l)}, L^{(l)}\}$  at the  $l$ th iteration

draw  $\Omega^{(l+1)}$  from  $p(\Omega \mid Y^{*(l)}, \vartheta^{(l)}, \pi^{(l)}, \Theta^{*(l)}, L^{(l)}, \beta^{(l)}, Y)$ ,

draw  $(\vartheta^{(l+1)}, Y^{*(l+1)})$  from  $p(\vartheta, Y^* \mid \Omega^{(l+1)}, \pi^{(l)}, \Theta^{*(l)}, L^{(l)}, \beta^{(l)}, Y)$ ,

draw  $(\pi^{(l+1)}, \Theta^{*(l+1)})$  from  $p(\pi, \Theta^* \mid \Omega^{(l+1)}, Y^{*(l+1)}, \vartheta^{(l+1)}, L^{(l)}, \beta^{(l)}, Y)$ ,

draw  $L^{(l+1)}$  from  $p(L \mid \Omega^{(l+1)}, Y^{*(l+1)}, \vartheta^{(l+1)}, \pi^{(l+1)}, \Theta^{*(l+1)}, \beta^{(l)}, Y)$ ,

draw  $\beta^{(l+1)}$  from  $p(\beta \mid \Omega^{(l+1)}, Y^{*(l+1)}, \vartheta^{(l+1)}, \pi^{(l+1)}, \Theta^{*(l+1)}, L^{(l+1)}, Y)$ ,

and form  $\{\Omega^{(l+1)}, Y^{*(l+1)}, \pi^{(l+1)}, \Theta^{*(l+1)}, L^{(l+1)}\}$ . It can be shown that as  $l$  tends to infinity, the empirical distribution of  $\{\Omega^{(l)}, Y^{*(l)}, \pi^{(l)}, \Theta^{*(l)}, L^{(l)}\}$  converges to  $p(\Omega, Y^*, \vartheta, \pi, \Theta^*, L, \beta \mid Y)$  at any geometrical rate. The full conditional distributions and the implementation of the above algorithm are given in the Appendix.

**3.2. Model Selection.** Model selection is an important issue in Bayesian semiparametric modeling for latent variable model since it is of practical interest to compare different modelings for factor analytic models. Formal Bayesian model selection is accomplished by comparing the marginal predictive distribution of data across models. Consider the problem of comparing competing models  $M_1$  and  $M_2$ . Let  $p(Y \mid M_1)$  and  $p(Y \mid M_2)$  denote the marginal density of data  $Y$  under  $M_1$  and  $M_2$ , respectively. A popular choice for selecting models is achieved via Bayes factor (BF) (e.g., [37–39]). However, in view of the fact that computing BF involves the high-dimensional density which is hard to estimate well, we

prefer comparing the following logarithm of pseudomarginal likelihood (LPML) [40, 41]:

$$\text{LPML}(Y) = \sum_{i=1}^n \log(\text{CPO}_i), \quad (16)$$

where  $\text{CPO}_i$  is known as the conditional predictive ordinate (CPO) defined as

$$\begin{aligned} \text{CPO}_i &= p(y_i \mid Y_{(i)}) = \left[ \int \frac{1}{p(y_i \mid Y_{(i)}, \vartheta, \Theta^*, \pi)} p(\vartheta, \Theta^*, \pi \mid Y) d\vartheta d\Theta^* d\pi \right]^{-1} \\ &= \left[ \int \frac{1}{p(y_i \mid \vartheta, \Theta^*, \pi)} p(\vartheta, \Theta^*, \pi \mid Y) d\vartheta d\Theta^* d\pi \right]^{-1}. \end{aligned} \quad (17)$$

Here,  $Y_{(i)}$  is the data set  $Y$  with  $y_i$  removed. Obviously, from (17), we can see that  $\text{CPO}_i$  is the marginal posterior predictive density of  $y_i$  given  $Y_{(i)}$  and can be interpreted as the height of this marginal density at  $y_i$ . Thus, small values of LPML imply that  $Y$  does not support the model.

Based on MCMC sample  $\{(\Theta^{*(t)}, \pi^{(t)}, L^{(t)}, \vartheta^{(t)}, \beta^{(t)}) : t = 1, \dots, T\}$  already available in the estimation, a consistent estimate for LPML can be obtained via ergodic average given by

$$\widehat{\text{LPML}}(Y) = -\sum_{i=1}^n \log \left[ \frac{1}{T} \sum_{t=1}^T \frac{1}{p(y_i \mid \vartheta^{(t)}, \Theta^{*(t)}, \pi^{(t)})} \right]. \quad (18)$$

It is noted that, under our proposed model,

$$\begin{aligned} p(y_i \mid \vartheta, \Theta^*, \pi) &= \int p(y_i \mid \omega_i, L_i, \vartheta, \Theta^*) \\ &\cdot p(\omega_i \mid L_i, \vartheta, \Theta^*) p(L_i \mid \pi) d\omega_i dL_i \end{aligned} \quad (19)$$

which is complicated due to the existence of  $\Omega$  and  $L$ . This can be solved by Monte Carlo approximation. Specifically, given the current values  $\{\vartheta^{(l)}, \Theta^{*(l)}, \pi^{(l)}\}$  at the  $l$ th iteration, we draw (i)  $L_i^{lh}$  from  $p(L_i \mid \pi^{(l)})$  and (ii)  $\omega_i^{lh}$  from  $p(\omega_i \mid L_i^{lh}, \vartheta^{(l)}, \Theta^{*(l)})$  for  $h = 1, \dots, H$  and then evaluate  $p(y_i \mid \vartheta^{(l)}, \Theta^{*(l)}, \pi^{(l)})$  at the observation  $y_i$  through

$$\begin{aligned} \widehat{p}(y_i \mid \vartheta^{(l)}, \Theta^{*(l)}, \pi^{(l)}) \\ \approx \frac{1}{H} \sum_{h=1}^H p(y_i \mid \omega_i^{lh}, L_i^{lh}, \vartheta^{(l)}, \Theta^{*(l)}). \end{aligned} \quad (20)$$

Obviously, the distributions involved in (i) and (ii) are standard and sampling is rather straightforward and fast.

## 4. A Simulation Study

In this section, a simulation study to evaluate the performance of the proposed procedure is conducted. The

goal is to assess the accuracy of estimates under parametric, partly exchangeable, and semiparametric modelings when data take on the multimodality or heterogeneity. We consider the situation in which each observed vector consists of three-dimensional continuous vector and three-dimensional ordinal vector with threshold values  $\tau_j = (-1.0^*, -0.6, 0.3, 1.0^*)^T$  ( $j = 1, 2, 3$ ). We generate  $Y$  by first generating  $Y^*$  with  $y_i^* = (x_i^T, u_i^T)^T$  from the mixture of two factor analytic models with weights 0.45 and 0.55 and then transforming  $u_i$  into  $z_i$  ( $s = 3$ ) via (1) to create the ordinal observations, where  $x_i$  represents a  $6 \times 1$  observed continuous random vector and  $u_i$  is a  $3 \times 1$  latent continuous random vector. Each component in the mixture model is specified through the following measurement model: for  $m = 1, 2$ ,

$$y_i^* = \mu^{(m)} + \Lambda \omega_i^{(m)} + \epsilon_i^{(m)}, \quad \epsilon_i^{(m)} \sim N(0, \Psi_\epsilon^{(m)}). \quad (21)$$

The parameters involved in the components of mixture model are taken as  $\mu^{(1)} = -1.5 \times \mathbf{1}_6$ ,  $\mu^{(2)} = 1.0 \times \mathbf{1}_6$ ,  $\Psi_\epsilon^{(1)} = 0.36I_6$ ,  $\Psi_\epsilon^{(2)} = I_6$ ,

$$\begin{aligned} \Lambda^T &= \begin{bmatrix} 1^* & 0.8 & 0.8 & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 1^* & 0.8 & 0.8 \end{bmatrix}, \\ \Phi^{(1)} &= \begin{pmatrix} 1 & -0.3 \\ -0.3 & 1.0 \end{pmatrix}, \\ \Phi^{(2)} &= \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1.0 \end{pmatrix}, \end{aligned} \quad (22)$$

in which  $\mathbf{1}_6$  is a  $6 \times 1$  vector with all elements equal to one and  $I_6$  is a  $6 \times 6$  identity matrix. The elements with asterisks involved in loading matrix  $\Lambda$  and threshold parameters  $\{\tau_j\}_{j=1}^3$  are treated as fixed for identifying model (see Section 2.1). Based on these settings, random sample with size 500 is generated and 100 replications are completed for each combination.

Prior inputs in the prior distributions involved in the parametric components (see (14)) are as follows:  $H_{\epsilon 0k}$  and  $H_{\zeta 0}$  are diagonal matrices with the diagonal elements 1.0, and elements in  $\{\Lambda_{0k}, \Pi_0\}$  are equal to the true values, while prior inputs in the prior distribution of superparameter  $\beta$  (see (15)) are  $\mu_0 = 0_9$ ,  $\Sigma_0 = 100I_9$ ,  $\kappa_1 = \kappa_2 = 0.001$ ,  $R_0^\phi = 0.01I_2$ ,  $\rho_0 = \rho_0^\phi = 10$ ,  $\alpha_{\epsilon 0k} = \beta_{\epsilon 0k} = 2.0$ , and  $\tau_1 = \tau_2 = 2.0$ . Note that these values ensure approaching noninformative priors.

A few test runs are conducted to explore the effect of truncated levels on the estimates of unknown parameters and the convergence of the blocked Gibbs sampler. We take  $G = 50, 60, 70, 80, 90, 100, 200$ , and 300 and calculate the total sum of the root mean square (RMS) of estimates (see below for details). The resulting values are 1.9830, 1.7382, 1.6582, 1.5548, 1.4194, 1.4128, 1.4108, and 1.4101, respectively. It can be seen that the total sum of the root mean square (RMS) becomes rather stable when  $G \geq 80$ . In the following analysis, we set  $G = 100$  in our data analysis. For the threshold parameters  $\{\tau_{jk} : j = 1, 2, 3, k = 2, 3\}$ , we choose  $\sigma_{MHjk}^2 = 0.002$  (see Appendix) in MH algorithm to produce

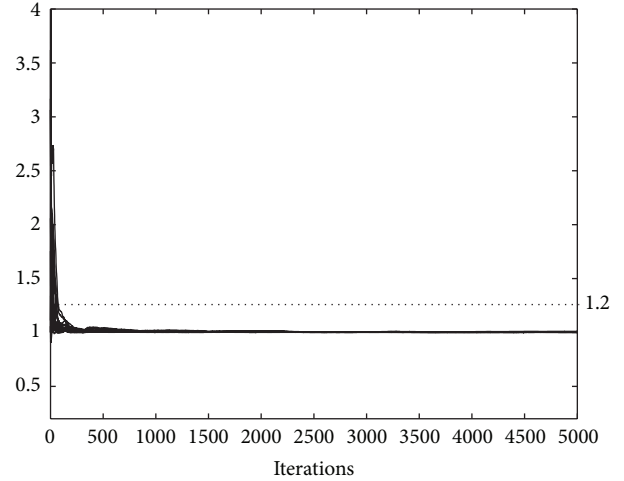


FIGURE 1: Plot of the values of EPSR of unknown parameters against the number of iterations under different starting values for the simulated data.

the acceptance rate about 0.40. Figure 1 gives the plots of EPSR (estimated potential scale reduction [42]) values of unknown free parameters in  $\Lambda$ ,  $\tau$ , and  $\alpha$  against iterations for three groups of different starting values. It can be seen that the estimates converge in less than 1000 iterations. To be conservative, in the following analysis, we collect 3000 observations after 2000 “burn-in”s deleted to take posterior analysis. We first consider the performance of the proposed LMPL in model comparison. We compare the proposed model with the parametric model (denoted by PARA) and the partly exchangeable model (denoted by PAEX), which approximately correspond to  $\alpha = +\infty$  and  $\alpha = 0$  under our proposal, respectively. The parametric model is defined by

$$\begin{aligned} (y_i^* | \omega_i, \theta) &\stackrel{\text{iid}}{\sim} N(\mu + \Lambda \omega_i, \Psi_\epsilon), \\ (\omega_i | \Phi) &\stackrel{\text{iid}}{\sim} N(0, \Phi). \end{aligned} \quad (23)$$

The priors of the parameters are given by  $p(\mu) \stackrel{D}{=} N(\mu_0, \Sigma_0)$ ,  $p(\Lambda, \Psi_\epsilon) \stackrel{D}{=} \prod_{k=1}^9 N(\Lambda_{0k}, H_{0k}) \cdot \text{Gamma}^{-1}(\alpha_{\epsilon 0k}, \beta_{\epsilon 0k})$ , and  $\Phi \sim \text{Wishart}^{-1}(10, 7.0I_2)$ .

The partly exchangeable model is given by

$$\begin{aligned} (y_i^* | \omega_i, \theta_i, \vartheta) &\stackrel{\text{iid}}{\sim} N(\mu_i + \Lambda \omega_i, \Psi_{\epsilon i}), \\ (\omega_i | \theta_i) &\stackrel{\text{iid}}{\sim} N(0, \Phi_i), \end{aligned} \quad (24)$$

where  $\theta_i = \{\mu_i, \Psi_{\epsilon i}, \Phi_i\}$  are i.i.d. with distribution  $F_0(\cdot | \nu, \Sigma, R)$  given in (5); the priors for the unknown parameter vector  $\Lambda$  and hyperparametric vector  $\{\nu, \Sigma, R\}$  are, respectively, given in (14) and (15).

Under the foregoing settings for the hyperparameters, observations obtained through the blocked Gibbs sampler are used to compute the values of LPML for each scenario across 100 replications. For the parametric and partly exchangeable model, computing values of LPML is very straightforward



and standard. For the semiparametric model, we draw 50 observations for approximating  $p(y_i | \vartheta, \Theta, \pi)$ . The values of LPML under parametric model, semiparametric model, and partly exchangeable model are, respectively,  $-6684.740$ ,  $-6255.553$ , and  $-8487.259$  with standard deviations  $62.509$ ,  $151.480$ , and  $147.742$ . Based on the LPML criteria, semiparametric model is selected, which is consistent with the fact that the true model takes on the multimodes. Moreover, according to our empirical results, the correct rates of LPML selecting the true model across 100 replications are about 0.93.

Table 1 gives the biases (BIAS), root mean squares (RMS), and standard deviations (SD) of estimates of unknown parameters across 100 replications under semiparametric models and parametric and partly exchangeable model, respectively. The measures BIAS, RMS, and SD are given as

$$\begin{aligned} \text{BIAS}(\hat{\vartheta}_j) &= (\bar{\vartheta}_j - \vartheta_{j0}), \quad \bar{\vartheta}_j = \frac{1}{S} \sum_{r=1}^S \hat{\vartheta}_j^{(r)}, \\ \text{RMS}(\hat{\vartheta}_j) &= \sqrt{\frac{1}{S} \sum_{r=1}^S (\hat{\vartheta}_j^{(r)} - \vartheta_{j0})^2}, \\ \text{SE}(\hat{\vartheta}_j) &= \sqrt{\frac{1}{S} \sum_{r=1}^S (\hat{\vartheta}_j^{(r)} - \bar{\vartheta}_j)^2}, \end{aligned} \quad (25)$$

where  $S$  is the number of replications. It can be seen that estimates obtained through the proposed approach are reasonably accurate. The values of  $\hat{\lambda}_{jk}$  under our approach are smaller than those under parametric and exchangeable modelings in terms of the absolute values of BIAS and RMS. The results show that ignoring heterogeneity among the data may lead to biased estimates and incorrect interpretation of the analyzed phenomena. This also reflects that the factor loadings  $\lambda_{jk}$  are not robust against the distributional deviations of inceptor, variance of unique errors, and covariances of latent factors.

Further simulation study is conducted to assess the performance of the proposed model and parametric model as well as the partly exchangeable model when data are generated from a single normal distribution. The population values of parameters are taken as  $\mu = 0_6$ ,  $\Psi_\epsilon = I_6$ , and

$$\Phi = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1.0 \end{pmatrix}. \quad (26)$$

The values of factor loadings and threshold points are the same as those in previous mixture model. As usual, we take  $G = 100$  for truncated levels. The sample size is set to 62 which is analogous to the real example. The inputs for superparameters involved in priors are set the same as that in mixture model. The results based on 100 replications are summarized in Table 2.

Based on Table 2, it can be found that the results obtained from our proposal are rather reasonable when compared to normal model, while partly exchangeable model gives serious biases. Moreover, we consider different inputs of superparameters in priors and find that the estimates are rather robust.

## 5. A Real Example

To illustrate the proposed procedure with a real example, a political-economic risk data set [43] was analyzed, which is adopted from Henisz's [44] political constraint index data set (POLCON), Marshall et al. [45] state failure problem sets (PITF), and Alvarez et al.'s [46] ACLP Political and Economic Database (ACLP). The data set is formed by the two economic indicators and three political variables from 62 countries. The first index is the log black market premium (BMP). This is a continuous variable which is usually used as a proxy for illegal economic activity. The second index is log real gross domestic product (GDP). It is used to measure the productivity of a country. The third variable is a measure of independence of the national judiciary. This is a binary variable: it takes 1 if the judiciary is judged to be independent and 0 otherwise. The next measurement, measuring the level of lack of expropriation risk threat (LE), is an ordered categorical variable coded with 0, 1, 2, 3, 4, and 5. The last variable is an expert judgment of measuring lack of corruption (LC). It is also an ordered categorical variable scaled with 0 to 5. The total sample size is 62 and the frequencies of each category occurring are equal to {34, 28}, {2, 6, 7, 19, 14, 14}, and {5, 11, 18, 11, 8, 9}, respectively. To unify scales of the continuous variables, the corresponding raw data were standardized.

Let  $y^T = (\log \text{BMP}, \log \text{GDP}, \text{IJ}, \text{LE}, \text{LC})$  be the vector of the observed variables. Based on the objective of this example, it is natural to group (i) {log BMP, log GDP} to an endogenous latent variable that can be interpreted as "economic factor,  $\xi$ " and (ii) {IJ, LE, LC} to an exogenous genotype latent variable that can be interpreted as "political factor,  $\eta$ ." Hence, the following loading matrix  $\Lambda$  in the measurement equation with  $\omega_i = (\eta_i, \xi_i)^T$  is considered:

$$\Lambda^T = \begin{bmatrix} 0^* & 0^* & 1^* & \lambda_{41} & \lambda_{51} \\ 1^* & \lambda_{22} & 0^* & 0^* & 0^* \end{bmatrix} \quad (27)$$

in which the ones and zeros are treated as known. Although other structures of  $\Lambda$  could be used, here we consider a nonoverlapped structure for clear interpretations of the latent variables:  $\lambda_{jk}$  measures the effect of  $\omega_k$  on the observed variable  $y_j$ . Since the third variable is binary and the last two variables are measured on a six-point scale with each involving six thresholds, for model identification, we fix  $\psi_{\epsilon 3} = 1$  and endpoints of thresholds  $\tau_{31}, \tau_{35}, \tau_{41}$ , and  $\tau_{45}$  at  $-1.8486$ ,  $0.7527$ ,  $-1.4007$ , and  $1.0574$ , respectively. These fixed threshold values were chosen via  $\tau_{jk} = \Phi^{-1}(\hat{p}_{jk})$ , where  $\hat{p}_{jk}$  are observed marginal proportions of the categories with  $z_j < k$ .

By primary data analysis, we find that the skewness and kurtosis of the first two variables are  $\{-0.1340, -0.4319\}$  and  $\{2.0892, 2.0958\}$ , respectively. We also evaluate the predictive density function for continuous variables. Figure 2 gives the contours of posterior predictive density of pair  $(y_1, y_2)$  under parametric model and semiparametric model  $M^e$  (see below) based on  $60 \times 60$  grids. It can be seen that the data for pair  $(y_1, y_2)$  are heavy-tailed and the predictive density under semiparametric model captures the high frequency

TABLE 1: Summary of the estimates under the parametric, partly exchangeable, and semiparametric approaches in analyzing simulated data: mixture data.

Para.	PARA			PAEX			SEMI		
	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD
$\lambda_{21}$	0.130	0.136	0.043	0.216	0.171	0.047	-0.022	0.066	0.070
$\lambda_{31}$	0.131	0.138	0.043	0.212	0.181	0.047	-0.026	0.080	0.070
$\lambda_{52}$	0.140	0.147	0.046	0.263	0.164	0.072	-0.015	0.093	0.092
$\lambda_{62}$	0.144	0.150	0.046	0.266	0.152	0.071	-0.012	0.098	0.092
$\tau_{12}$	0.060	0.091	0.040	-0.110	0.014	0.052	0.007	0.062	0.038
$\tau_{13}$	0.070	0.085	0.033	-0.216	0.050	0.060	0.007	0.067	0.040
$\tau_{22}$	0.070	0.098	0.040	-0.113	0.016	0.051	0.008	0.056	0.038
$\tau_{23}$	0.077	0.095	0.032	-0.186	0.038	0.059	0.013	0.062	0.042
$\tau_{32}$	0.062	0.092	0.040	-0.118	0.016	0.051	0.013	0.069	0.038
$\tau_{33}$	0.083	0.098	0.032	-0.184	0.037	0.059	0.018	0.070	0.042

TABLE 2: Summary of the estimates under the parametric, partly exchangeable, and semiparametric approaches in analyzing simulated data: normal data.

Para.	PARA			PAEX			SEMI		
	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD
$\lambda_{21}$	0.109	0.103	0.068	0.216	0.201	0.147	0.152	0.125	0.130
$\lambda_{31}$	-0.114	0.108	0.066	0.212	0.201	0.207	0.138	0.137	0.120
$\lambda_{52}$	-0.122	0.116	0.101	0.253	0.134	0.172	0.146	0.139	0.141
$\lambda_{62}$	-0.123	0.114	0.102	0.366	0.136	0.271	0.144	0.118	0.176
$\tau_{12}$	-0.012	0.004	0.052	-0.151	0.036	0.126	0.005	0.002	0.052
$\tau_{13}$	-0.018	0.002	0.056	-0.153	0.041	0.126	-0.013	0.003	0.060
$\tau_{22}$	-0.001	0.003	0.046	-0.110	0.014	0.052	-0.035	0.004	0.045
$\tau_{23}$	0.016	0.003	0.051	-0.216	0.050	0.060	-0.022	0.004	0.055
$\tau_{32}$	-0.001	0.003	0.046	-0.113	0.016	0.051	-0.001	0.002	0.048
$\tau_{33}$	-0.018	0.002	0.051	-0.186	0.038	0.059	-0.012	0.003	0.058

region successfully while parametric model fails. For model comparison, we consider the following competing models:

$$\begin{aligned}
 M^\mu: y_i^* &= \mu_i + \Lambda\omega_i + \epsilon_i, \\
 &\epsilon_i \sim N(0, \Psi_\epsilon), \omega_i \sim N(0, \phi); \\
 M^\epsilon: y_i^* &= \mu + \Lambda\omega_i + \epsilon_i, \\
 &\epsilon_i \sim N(0, \Psi_{\epsilon i}), \omega_i \sim N(0, \phi); \\
 M^{\mu\epsilon}: y_i^* &= \mu_i + \Lambda\omega_i + \epsilon_i, \\
 &\epsilon_i \sim N(0, \Psi_{\epsilon i}), \omega_i \sim N(0, \phi); \\
 M^{\mu\phi}: y_i^* &= \mu_i + \Lambda\omega_i + \epsilon_i, \\
 &\epsilon_i \sim N(0, \Psi_\epsilon), \omega_i \sim N(0, \phi_i); \\
 M^{\epsilon\phi}: y_i^* &= \mu + \Lambda\omega_i + \epsilon_i, \\
 &\epsilon_i \sim N(0, \Psi_{\epsilon i}), \omega_i \sim N(0, \phi_i); \\
 M^{\mu\epsilon\phi}: y_i^* &= \mu_i + \Lambda\omega_i + \epsilon_i, \\
 &\epsilon_i \sim N(0, \Psi_{\epsilon i}), \omega_i \sim N(0, \phi_i).
 \end{aligned}
 \tag{28}$$

The following two types of prior inputs are, respectively, used for the hyperparameters involved in the parametric components and semiparametric components: (I)  $\Lambda_{\epsilon 0k} = \tilde{\Lambda}_{\epsilon k}, H_{\epsilon 0k} = I_2, \gamma_0 = \tilde{\gamma}, H_{\zeta 0} = 1, \mu_0 = \tilde{\mu}, \Sigma_0 = \text{diag}\{S\}, \alpha_{\epsilon 0k} = 9.0, \beta_{\epsilon 0k} = (\alpha_{\epsilon 0k} - 1)\tilde{\psi}_{\epsilon k}, \rho_0 = \rho_0^\phi = 20, R_0^{\phi-1} = (\rho_0 - 2)\tilde{\phi}, \kappa_1 = \kappa_2 = 8.0,$  and  $\tau_1 = \tau_2 = 8.0,$  where  $\tilde{\theta}$  denotes the maximum likelihood estimates of  $\theta$  under parametric model from analysis of a ‘‘control-group’’ sample and  $S$  is the polychoric correlation matrix obtained on the basis of single confirmatory factor analysis model; (II)  $\lambda_{\epsilon 0jk} = 0, H_{\epsilon 0k} = 0.01I_2, \gamma_0 = 0, H_{\zeta 0} = 0.01, \mu_0 = 0_{5 \times 1}, \Sigma_0 = 0.01I_5, \alpha_{\epsilon 0k} = \beta_{\epsilon 0k} = 2.0, \rho_0 = \rho_0^\phi = 10, R_0^{-1} = \rho_0 - 2, \kappa_1 = \kappa_2 = 0.01,$  and  $\tau_1 = \tau_2 = 0.01.$  Note that prior (I) gives more information than prior (II) since it partly takes advantage of information from sample.

The proposed Bayesian semiparametric approach with  $G = 50$  was applied to calculate the values of CPO and LPML. We draw 100,000 effective observations from the corresponding posteriors via the blocked Gibbs sampler and divide them into 100 batches equally. Table 3 gives the means and standard deviations of LPML under priors (I) and (II). The following facts can be found. (i) The values of LPML under prior (I) are larger than those under prior (II). This indicates that the LPML tends to choose the model with

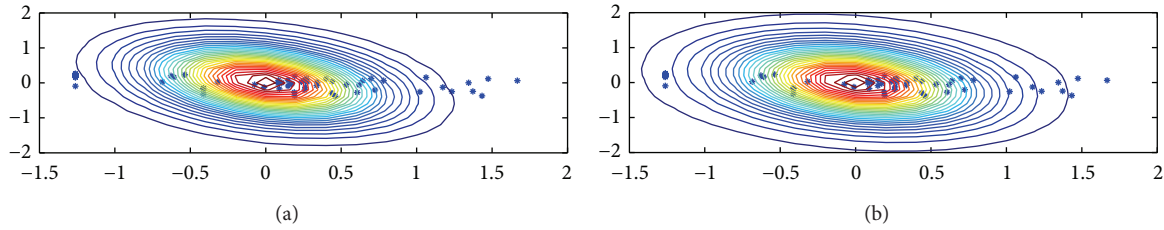


FIGURE 2: Plot of contours of posterior predictive density of pair  $(y_1, y_2)$  under parametric model and semiparametric model  $M^\epsilon$ : (a) corresponds to parametric model and (b) corresponds to semiparametric model with  $G = 50$ .

TABLE 3: Mean and standard deviation (SD) of LPML in the political and economic risk data.

Model	LPML			
	BAY I		BAY II	
	Mean	SD	Mean	SD
$M^\mu$	-253.0577	60.7693	-274.9704	14.7176
$M^\epsilon$	-176.5289	110.0227	-188.8361	175.3741
$M^{\mu\epsilon}$	-193.2270	81.3209	-209.0063	77.0664
$M^{\mu\phi}$	-235.7465	15.3588	-235.7465	11.6911
$M^{\epsilon\phi}$	-267.4536	10.2084	-271.2875	10.4206
$M^{\mu\epsilon\phi}$	-255.2335	27.2208	-258.7124	15.8416

informative prior. (ii)  $M^\epsilon$  give the largest value. Among the posited models,  $M^\epsilon$  is selected. We also compute the values of LPML for parametric model. They are  $-287.4262$  and  $-288.6033$  under priors (I) and (II) with standard deviations  $6.287$  and  $5.065$ , respectively. Therefore, the data support the semiparametric model instead of parametric model.

Table 4 presents the estimates of factor loading  $\lambda_{jk}$  as well as their standard deviations with semiparametric and parametric model under prior (I). The factor loading estimates  $\hat{\lambda}_{kj}$  in the measurement equation can be interpreted according to a standard confirmatory factor analysis model. The difference between the two approaches is obvious: the estimates of  $\lambda_{41}$  and  $\lambda_{51}$  under parametric model are only half of those under semiparametric model. Moreover, the standard deviations of estimates with parametric method are uniformly larger than that of semiparametric model. Since we identify illegal economic activity log PCR with economic factor  $\xi$  ( $\lambda_{12}^* = 1$ ) and independent of judiciary with political factor  $\eta$  ( $\lambda_{31}^* = 1$ ), respectively, the level of economic factor has a negative effect on real gross domestic product, while the level of political factor has positive effect on lack of expropriation risk threat and lack of corruption. The estimate  $\hat{\lambda}_{22} = -0.123$  indicates that a one-unit increase in the level of economic factor leads to 0.123-unit decrease in the magnitude of gross domestic product. The interpretation of  $\hat{\lambda}_{41}$  and  $\hat{\lambda}_{51}$  is similar. The differences of estimates between parametric and semiparametric methods illustrate the effects of heavy tails of the data on the estimates.

### 6. Concluding Remarks

Parametric modeling for latent variable model with mixed data structure has long dominated Bayesian inference work,

TABLE 4: Estimates and standard errors estimates of the parameters in analysis of political and economic risk data.

Parameter	Parametric model		$M^\epsilon$ model	
	Est.	SD	Est.	SD
$\lambda_{22}$	-0.155	0.083	-0.123	0.077
$\lambda_{41}$	0.418	0.104	0.846	0.088
$\lambda_{51}$	0.367	0.090	0.754	0.066
$\tau_{42}$	-1.336	0.157	-1.340	0.055
$\tau_{43}$	-0.905	0.167	-0.898	0.061
$\tau_{44}$	-0.008	0.149	0.004	0.055
$\tau_{52}$	-0.794	0.142	-0.787	0.046
$\tau_{53}$	0.001	0.155	0.002	0.053
$\tau_{54}$	0.566	0.136	0.567	0.035

typically developed within the standard exponential family. Such modeling is often confused with handling the multimodal and unknown heterogeneous problems. In dealing with multimodality or increased heterogeneity in data, one naturally resorts to the finite mixture model [47, 48] which is more flexible and feasible to implement due to advances in simulation-based model fitting.

Rather than handling the large number of parameters resulting from the finite mixture models with a large number of components, we consider, in this paper, the finite-dimensional Dirichlet process mixture model for latent variable model with continuous and ordinal responses. The core of our proposal is to model the mean vector and/or variance-covariance parameters of unique errors and latent variables into the finite-dimensional stick-breaking priors. This will help to reveal the local dependence structure such as classification groups and clustering among the data. The blocked Gibbs sampler developed by Ishwaran and Zarepour [31], which takes advantage of the block updating and accelerates mixing in Gibbs sampling, is adapted here to cope with the posterior inference.

The proposed methodologies in this paper can be applied to more general latent variable models that include the multilevel SEMs [49] and longitudinal latent trait models [5] with discrete variables.

### Appendix

#### Full Conditional Distributions

(1) Full Conditional Distribution  $p(\vartheta, Y^* \mid \Omega, \pi, \Theta^*, L, \beta, Y)$ . To draw  $(\vartheta, Y^*)$  from  $p(\vartheta, Y^* \mid \Omega, \pi, \Theta^*, L, \beta, Y)$ ,



we implement it by (i) drawing  $(\tau, Y^*)$  from  $p(\tau, Y^* \mid \Omega, \Lambda, \pi, \Theta^*, L, \beta, Y)$  and (ii) drawing  $\Lambda$  from  $p(\Lambda \mid \Omega, \tau, Y^*, \pi, \Theta^*, L, \beta, Y)$ . The underlying reason is that drawing  $(\tau, Y^*)$  from the joint conditional distribution as proposed here is more efficient than drawing  $\tau$  and  $Y^*$  separately from the corresponding marginal conditional distribution (see Liu [50], Nandram and Chen [51], and Song and Lee [6]).

It can be shown that  $p(\tau, Y^* \mid \Omega, \Lambda, \pi, \Theta^*, L, \beta, Y)$ , not involving  $\pi$  and  $\beta$ , is given by

$$\begin{aligned} p(\tau, Y^* \mid \Omega, \Lambda, \Theta^*, L, Y) &= p(\tau \mid \Omega, \Lambda, \Theta^*, L, Y) \\ &\cdot p(Y^* \mid \tau, \Omega, \Lambda, \Theta^*, L, Y) \\ &= \prod_{j=1}^s p(\tau_j \mid \Omega, \Lambda, \Theta_{(j)}, Y_{(j)}) \\ &\cdot p(Y_{(j)}^* \mid \tau_j, \Omega, \Lambda, \Theta_{(j)}, Y_{(j)}), \end{aligned} \tag{A.1}$$

where  $Y_{(j)}^* = \{y_{ij}^* : i = 1, \dots, n\}$ ,  $\Theta_{(j)} = \{\theta_{ij} : i = 1, \dots, n\}$ , and  $Y_{(j)} = \{y_{ij} : i = 1, \dots, n\}$ . Further,

$$\begin{aligned} p(\tau_j \mid \Omega, \Lambda, \Theta_{(j)}, Y_{(j)}) \\ \propto \prod_{k=1}^{b_j-1} \prod_{\{i: y_{ij}=k\}} \{\Phi(\psi_{\epsilon ij}^{-1/2}(\tau_{j,k+1} - \mu_{ij} - \Lambda_j^T \omega_i))\} \end{aligned}$$

$$\begin{aligned} &- \Phi(\psi_{\epsilon ij}^{-1/2}(\tau_{j,k} - \mu_{ij} - \Lambda_j^T \omega_i))\} I\{\tau_{j,k} < \tau_{j,k+1}\}, \\ p(Y_{(j)}^* \mid \tau_j, \Omega, \Lambda, \Theta_{(j)}, Y_{(j)}) &\propto \prod_{i=1}^n N(y_{ij}^* \mid \mu_{ij} \\ &+ \Lambda_j^T \omega_i, \psi_{\epsilon ij}) I\{\tau_{j,y_{ij}} < y_{ij}^* \leq \tau_{j,y_{ij}+1}\}, \end{aligned} \tag{A.2}$$

where  $\Phi(\cdot)$  is the cumulative distribution function of  $N(0, 1)$ . It is difficult to sample  $\tau_j$  from  $p(\tau_j \mid \Omega, \Lambda, \Theta_{(j)}, Y_{(j)})$  since this target distribution is nonstandard. We follow Cowles' routines [52] and use Metropolis-Hasting (MH) algorithm to sample observations from this complex conditional distribution. Specifically, given the current values  $\tau_j^{(l)} = (\tau_{j,2}^{(l)}, \dots, \tau_{j,b_j-1}^{(l)})^T$  at the  $l$ th iteration, generate a candidate vector  $\tau_j^* = (\tau_{j,2}^*, \dots, \tau_{j,b_j-1}^*)^T$  from the following truncated normal distribution:

$$\begin{aligned} \tau_{j,k}^* \sim N(\tau_{j,k}^{(l)}, \sigma_{MHjk}^2) I\{(\tau_{j,k-1}^{(l)}, \tau_{j,k+1}^{(l)})\}, \\ \text{for } k = 2, \dots, b_j - 1. \end{aligned} \tag{A.3}$$

Accept this candidate  $\tau_j^*$  as  $\tau_j^{(l+1)}$  with the probability  $\min\{1, R_j\}$ , where

$$\begin{aligned} R_j &= \prod_{k=2}^{b_j-1} \frac{\Phi(\sigma_{MHjk}^{-1}[\tau_{j,k+1}^{(l)} - \tau_{j,k}^{(l)}]) - \Phi(\sigma_{MHjk}^{-1}[\tau_{j,k-1}^* - \tau_{j,k}^{(l)}])}{\Phi(\sigma_{MHjk}^{-1}[\tau_{j,k+1}^* - \tau_{j,k}^*]) - \Phi(\sigma_{MHjk}^{-1}[\tau_{j,k-1}^{(l)} - \tau_{j,k}^*])} \\ &\times \prod_{i=1}^n \frac{\Phi(\psi_{\epsilon ij}^{-1/2}(\tau_{j,y_{ij}+1}^* - \mu_{ij} - \Lambda_j^T \omega_i)) - \Phi(\psi_{\epsilon ij}^{-1/2}(\tau_{j,y_{ij}}^* - \mu_{ij} - \Lambda_j^T \omega_i))}{\Phi(\psi_{\epsilon ij}^{-1/2}(\tau_{j,y_{ij}+1}^{(l)} - \mu_{ij} - \Lambda_j^T \omega_i)) - \Phi(\psi_{\epsilon ij}^{-1/2}(\tau_{j,y_{ij}}^{(l)} - \mu_{ij} - \Lambda_j^T \omega_i))}. \end{aligned} \tag{A.4}$$

As pointed out by Cowles (1996) [52], the quantities  $\sigma_{MHjk}^2$  should be chosen carefully such that the average acceptance probability is about 0.30 or more.

For  $p(\Lambda, \Pi \mid \Omega, \tau, Y^*, \pi, \Theta^*, L, \beta, Y)$ , without loss of generality, we assume that the elements in  $\Lambda$  are all free. Let  $y_{ij}^{**} = y_{ij}^* - \mu_j$ . Under the prior distributions given in (14), we have

$$p(\Lambda \mid \Omega, Y^*, \Theta^*, L) \stackrel{D}{=} \prod_{k=1}^p N(m_{\epsilon k}, \Sigma_{\epsilon k}), \tag{A.5}$$

in which

$$\begin{aligned} m_{\epsilon k} &= \Sigma_{\epsilon k} \left( H_{\epsilon 0k}^{-1} \Lambda_{0k} + \sum_{i=1}^n \frac{\omega_i y_{ik}^{**}}{\psi_{\epsilon ik}} \right), \\ \Sigma_{\epsilon k} &= \left( \sum_{i=1}^n \psi_{\epsilon ik}^{-1} \omega_i \omega_i^T + H_{\epsilon 0k}^{-1} \right)^{-1}. \end{aligned} \tag{A.6}$$

(2) Full Conditional Distribution  $p(\Omega \mid \tau, Y^*, \theta, \pi, \Theta^*, L, \beta, Y)$ . It can be shown that the conditional distribution of  $\Omega$  is given by

$$\begin{aligned} p(\Omega \mid \tau, Y^*, \theta, \pi, \Theta^*, L, \beta, Y) \\ = \prod_{i=1}^n p(\omega_i \mid \theta, \theta_i, y_i^*), \\ [\omega_i \mid \theta, \theta_i, y_i^*] \stackrel{D}{=} N(\Sigma_{\omega i} \Lambda^T \Psi_{\epsilon i}^{-1} (y_i^* - \mu_i), \Sigma_{\omega i}), \end{aligned} \tag{A.7}$$

where  $\Sigma_{\omega i} = (\Lambda^T \Psi_{\epsilon i}^{-1} \Lambda + \Phi_i^{-1})^{-1}$ .

(3) The Full Conditional Distribution  $p(\pi, \Theta^* \mid \Omega, Y^*, \vartheta, L, \beta, Y)$ . It is clear that

$$\begin{aligned} p(\pi, \Theta^* \mid \Omega, Y^*, \vartheta, L, \beta, Y) \\ = p(\pi \mid \Theta^*, \Omega, Y^*, \vartheta, L, \beta, Y) \\ \cdot p(\Theta^* \mid \pi, \Omega, Y^*, \vartheta, L, \beta, Y). \end{aligned} \tag{A.8}$$

Let  $m_k = \#\{i : L_i = k\}$  be the number of  $L_i$  equal to  $k$ , for  $k = 1, \dots, G - 1$ . It can be shown that  $p(\pi | \Theta^*, \Omega, Y^*, \vartheta, L, \beta, Y) = p(\pi | L, \alpha)$  is a generalized Dirichlet distribution,  $\mathcal{GD}(a_1^*, b_1^*, \dots, a_{G-1}^*, b_{G-1}^*)$  with  $a_k^* = 1 + m_k$ ,  $b_k^* = \alpha + \sum_{j=k+1}^G m_j$  ( $k = 1, \dots, G - 1$ ), which is constructed by

$$\begin{aligned} \pi_1 &= V_1^*, \\ \pi_k &= V_k^* \prod_{j=1}^{k-1} (1 - V_j^*) \quad (k = 2, \dots, G - 1), \end{aligned} \tag{A.9}$$

where  $V_j^* \stackrel{\text{ind}}{\sim} \text{Beta}(a_j^*, b_j^*)$ .

For  $p(\Theta^* | \pi, \Omega, Y^*, \vartheta, L, \beta, Y) = p(\Theta^* | \Omega, Y^*, \vartheta, L, \beta)$ , let  $L^* = \{L_1^*, \dots, L_m^*\}$  be the unique set of  $L$ ,  $\Theta_{L^*}^* = \{\theta_{L_1^*}^*, \dots, \theta_{L_m^*}^*\}$ , and  $\Theta_{(-L^*)}^*$  corresponding to those values in  $\Theta^*$  with  $\Theta_{L^*}^*$  excluded. Then,

$$\begin{aligned} p(\Theta^* | \Omega, Y^*, \vartheta, L, \beta) &= p(\Theta_{(-L^*)}^* | \Omega, Y^*, \vartheta, L, \beta) \\ &\cdot p(\Theta_{L^*}^* | \Theta_{(-L^*)}^*, \Omega, Y^*, \vartheta, L, \beta) \\ &= p(\Theta_{(-L^*)}^* | \beta) p(\Theta_{L^*}^* | \Omega, Y^*, \vartheta, L, \beta). \end{aligned} \tag{A.10}$$

Let  $\mu^* = \{\mu_j^* : j = 1, \dots, G\}$ ,  $\Psi_\epsilon^* = \{\Psi_{\epsilon_j}^* : j = 1, \dots, G\}$ , and  $\Phi^* = \{\Phi_j^* : j = 1, \dots, G\}$ , and note that  $\Theta^* = \{\mu^*, \Psi_\epsilon^*, \Phi^*\}$ . The components of  $\{\mu_{(-L^*)}^*, \Psi_{\epsilon(-L^*)}^*, \Phi_{(-L^*)}^*\}$  are easy to sample based on (14). Further,

$$\begin{aligned} p(\Theta_{L^*}^* | \Omega, Y^*, \vartheta, L, \beta) \\ = \prod_{j=1}^m p(\theta_{L_j^*}^* | \Omega, Y^*, \vartheta, L, \beta) \end{aligned} \tag{A.11}$$

which can be implemented by drawing: for  $l \in L^*$

$$\begin{aligned} p(\mu_l^* | \Psi_{\epsilon l}^*, \Psi_{\zeta l}^*, \Phi_l^*, \Omega, Y^*, \vartheta, L, \beta) &\sim N(\mu_l^*, \Sigma_l^*), \\ p(\Psi_{\epsilon l}^{*-1} | \mu_l^*, \Psi_{\zeta l}^*, \Phi_l^*, \Omega, Y^*, \vartheta, L, \beta) \\ &\sim \prod_{k=1}^p \text{Gamma}(\alpha_{\epsilon l k}^*, \beta_{\epsilon l k}^*), \\ p(\Phi_l^{*-1} | \mu_l^*, \Psi_{\epsilon l}^*, \Psi_{\zeta l}^*, \Omega, Y^*, \vartheta, L, \beta) \\ &\sim \text{Wishart}(\rho_{l0}^*, R_{l0}^*) \end{aligned} \tag{A.12}$$

in which

$$\begin{aligned} \mu_l^* &= \Sigma_l^* \left( \Sigma_l^{-1} \nu + \left( \Psi_{\epsilon l}^{*-1} \sum_{\{i:L_i=l\}} (y_i^* - \Lambda \omega_i) \right) \right), \\ \Sigma_l^* &= \left[ \Sigma_\nu^{-1} + m_l \Psi_{\epsilon l}^{*-1} \right]^{-1}, \\ \alpha_{\epsilon l k}^* &= \alpha_{\epsilon 0k} + m_l, \\ \beta_{\epsilon l k}^* &= \beta_{\epsilon 0k} + 2^{-1} \sum_{\{i:L_i=l\}} (y_{ik}^* - \mu_{ik}^* - \Lambda_k^T \omega_i)^2, \end{aligned}$$

$$\rho_{l0}^* = \rho_0^\phi + m_l,$$

$$R_{l0}^* = \left( R^{\phi-1} + \sum_{\{i:L_i=l\}} \omega_i \omega_i^T \right)^{-1}. \tag{A.13}$$

(4) Full Conditional Distribution  $p(L | \Omega, \tau, Y^*, \vartheta, \pi, Z, Y)$ . It can be shown that

$$\begin{aligned} p(L | \Omega, Y^*, \vartheta, \pi, \Theta^*, Y) \\ = \prod_{i=1}^n p(L_i | \Omega, Y^*, \vartheta, \pi, \Theta^*, Y), \\ [L_i = \cdot | \Omega, \tau, Y^*, \vartheta, \pi, \Theta^*, Y] \stackrel{\text{iid}}{\sim} \sum_{k=1}^G \pi_{ik}^* \delta_k(\cdot), \end{aligned} \tag{A.14}$$

where  $\pi_{ik}^* = c_i \pi_k p(y_i^* | \omega_i, \theta_k^*, \vartheta) p(\omega_i | \theta_k^*)$  and  $c_i$  is a normalized constant such that  $\sum_{k=1}^G \pi_{ik}^* = 1.0$ .

(5) Full Conditional Distribution  $p(\beta | \Omega, \tau, Y^*, \vartheta, \pi, Z, L, Y)$ . Based on the priors given in (15), the full conditional distributions for components of hyperparameters  $\beta$  are given as follows:

$$\begin{aligned} [\nu | \Theta^*, \Sigma_\nu] &\sim N(m_\nu, A_\nu), \\ [\Sigma_\nu, \nu, \Theta^*] &\sim \prod_{j=1}^p \text{Gamma}^{-1} \left( \kappa_1 + 0.5G, \kappa_2 \right. \\ &\quad \left. + 0.5 \sum_{k=1}^G (\mu_{kj}^* - \mu_j)^2 \right), \end{aligned} \tag{A.15}$$

$$\begin{aligned} [R^{-1} | \Theta^*] &\sim \text{Wishart} \left( G\rho_0 + \rho_0^\phi, \right. \\ &\quad \left. \left( \sum_{k=1}^G \Phi_k^{*-1} + R_0^\phi \right)^{-1} \right), \end{aligned}$$

$$[\alpha | \pi] \sim \text{Gamma}(\tau_1 + G - 1, \tau_2 - \log \pi_G),$$

where  $m_\nu = A_\nu \{\Sigma_0^{-1} \mu_0 + \Sigma_\nu^{-1} \sum_{k=1}^G \mu_k^*\}$  and  $A_\nu = (G\Sigma_\nu^{-1} + \Sigma_0^{-1})^{-1}$ .

### Competing Interests

The authors declare that they have no competing interests.

### Acknowledgments

The work described in this paper was supported by the National Natural Science Fund (11471161), Nanjing Forestry University Foundation (163101004), and Technological Innovation Item of Personnel Division (013101001).

## References

- [1] B. Muthén, *LISCOMP: Analysis of Linear Statistical Equation with a Comprehensive Measurement Model*, Scientific Software, Mooresville, NC, USA, 1987.
- [2] J.-Q. Shi and S.-Y. Lee, “Bayesian sampling-based approach for factor analysis models with continuous and polytomous data,” *British Journal of Mathematical and Statistical Psychology*, vol. 51, no. 2, pp. 233–252, 1998.
- [3] J.-Q. Shi and S.-Y. Lee, “Latent variable models with mixed continuous and polytomous data,” *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, vol. 62, no. 1, pp. 77–87, 2000.
- [4] D. B. Dunson, “Bayesian latent variable models for clustered mixed outcomes,” *Journal of the Royal Statistical Society—Series B: Statistical Methodology*, vol. 62, no. 2, pp. 355–366, 2000.
- [5] D. B. Dunson, “Dynamic latent trait models for multidimensional longitudinal data,” *Journal of the American Statistical Association*, vol. 98, no. 463, pp. 555–563, 2003.
- [6] X.-Y. Song and S.-Y. Lee, “Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data,” *British Journal of Mathematical and Statistical Psychology*, vol. 57, no. 1, pp. 29–52, 2004.
- [7] Y. Kano, M. Berkane, and P. M. Bentler, “Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations,” *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 135–143, 1993.
- [8] Y. Kano, “Consistency property of elliptical probability density functions,” *Journal of Multivariate Analysis*, vol. 51, no. 1, pp. 139–147, 1994.
- [9] S.-Y. Lee and Y.-M. Xia, “Maximum likelihood methods in treating outliers and symmetrically heavy-tailed distributions for nonlinear structural equation models with missing data,” *Psychometrika*, vol. 71, no. 3, pp. 565–585, 2006.
- [10] S.-Y. Lee and Y.-M. Xia, “A robust Bayesian approach for structural equation models with missing data,” *Psychometrika*, vol. 73, no. 3, pp. 343–364, 2008.
- [11] Y.-M. Xia, X.-Y. Song, and S.-Y. Lee, “Robust model fitting for the non linear structural equation model under normal theory,” *British Journal of Mathematical and Statistical Psychology*, vol. 62, no. 3, pp. 529–568, 2009.
- [12] Y.-M. Xia and Y.-A. Liu, “Robust Bayesian analysis and its applications for a factor analytic model with normal scale mixing,” *Chinese Journal of Applied Probability and Statistics*, vol. 30, no. 4, pp. 423–438, 2014.
- [13] A. Kottas, P. Müller, and F. Quintana, “Nonparametric Bayesian modeling for multivariate ordinal data,” *Journal of Computational & Graphical Statistics*, vol. 14, no. 3, pp. 610–625, 2005.
- [14] J. H. Albert and S. Chib, “Bayesian analysis of binary and polytomous response data,” *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 669–679, 1993.
- [15] M. K. Cowles, B. P. Carlin, and J. E. Connett, “Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness,” *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 86–98, 1996.
- [16] M.-H. Chen and D. K. Dey, “Bayesian analysis for correlated ordinal data models,” in *Generalized Linear Models: A Bayesian Perspectives*, D. K. Dey, S. Ghosh, and B. K. Mallick, Eds., pp. 133–157, Marcel Dekker, New York, NY, USA, 2000.
- [17] S.-Y. Lee, B. Lu, and X.-Y. Song, “Semiparametric Bayesian analysis of structural equation models with fixed covariates,” *Statistics in Medicine*, vol. 27, no. 13, pp. 2341–2360, 2008.
- [18] H. Ishwaran and L. F. James, “Gibbs sampling methods for stick-breaking priors,” *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 161–173, 2001.
- [19] X.-Y. Song, Y.-M. Xia, and S.-Y. Lee, “Bayesian semiparametric analysis of structural equation models with mixed continuous and unordered categorical variables,” *Statistics in Medicine*, vol. 28, no. 17, pp. 2253–2276, 2009.
- [20] X.-Y. Song, Y.-M. Xia, J.-H. Pan, and S.-Y. Lee, “Model comparison of bayesian semiparametric and parametric structural equation models,” *Structural Equation Modeling—A Multidisciplinary Journal*, vol. 18, no. 1, pp. 55–72, 2011.
- [21] Y.-M. Xia and A.-Y. Liu, “Bayesian semiparametric analysis for generalized linear latent variable model,” in *The Proceedings of 2010 International Conference on Probability and Statistics of the International Institute for General Systems Studies*, Y. Jiang and G. Z. Wang, Eds., vol. 1 of *Advances on Probability and Statistics*, pp. 308–312, 2010.
- [22] Y.-M. Xia and Y.-A. Liu, “Bayesian semiparametric analysis and model comparison for confirmatory factor model,” *Chinese Journal of Applied Probability and Statistics*, In press.
- [23] Y.-M. Xia and J.-W. Gou, “Semiparametric Bayesian analysis for factor analysis model mixed with hidden Markov model,” *Applied Mathematics*, vol. 30, pp. 17–30, 2015.
- [24] S.-Y. Lee and H.-T. Zhu, “Statistical analysis of nonlinear structural equation models with continuous and polytomous data,” *British Journal of Mathematical and Statistical Psychology*, vol. 53, no. 2, pp. 209–232, 2000.
- [25] T. S. Ferguson, “A Bayesian analysis of some nonparametric problems,” *The Annals of Statistics*, vol. 1, pp. 209–230, 1973.
- [26] C. E. Antoniak, “Mixtures of dirichlet processes with applications to bayesian nonparametric problems,” *The Annals of Statistics*, vol. 2, no. 6, pp. 1152–1174, 1974.
- [27] A. Y. Lo, “On a class of Bayesian nonparametric estimates. I. Density estimates,” *The Annals of Statistics*, vol. 12, no. 1, pp. 351–357, 1984.
- [28] M. D. Escobar, “Estimating normal means with a Dirichlet process prior,” *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 268–277, 1994.
- [29] M. D. Escobar and M. West, “Bayesian density estimation and inference using mixtures,” *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 577–588, 1995.
- [30] P. Müller and F. A. Quintana, “Nonparametric Bayesian data analysis,” *Statistical Science*, vol. 19, no. 1, pp. 95–110, 2004.
- [31] H. Ishwaran and M. Zarepour, “Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models,” *Biometrika*, vol. 87, no. 2, pp. 371–390, 2000.
- [32] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, New York, NY, USA, 1971.
- [33] M. A. Tanner and W. H. Wong, “The calculation of posterior distributions by data augmentation,” *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–550, 1987.
- [34] S. Geman and D. Geman, “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [35] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [36] W. K. Hastings, “Monte carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

- [37] J. Albert and S. Chib, "Bayesian tests and model diagnostics in conditionally independent hierarchical models," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 916–925, 1997.
- [38] R. E. Kass and A. E. Raftery, "Bayes factors," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995.
- [39] S. Basu and S. Chib, "Marginal likelihood and Bayes factors for Dirichlet process mixture models," *Journal of the American Statistical Association*, vol. 98, no. 461, pp. 224–235, 2003.
- [40] S. Geisser and W. F. Eddy, "A predictive approach to model selection," *Journal of the American Statistical Association*, vol. 74, no. 365, pp. 153–160, 1979.
- [41] S. Mukhopadhyay and A. E. Gelfand, "Dirichlet process mixed Generalized linear models," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 633–639, 1997.
- [42] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, pp. 457–472, 1992.
- [43] K. M. Quinn, "Bayesian factor analysis for mixed ordinal and continuous responses," *Political Analysis*, vol. 12, no. 4, pp. 338–353, 2004.
- [44] W. J. Henisz, *The Political Constraint Index (POLCON) Dataset*, 2002, <http://www-management.wharton.upenn.edu/henisz/POLCON/ContactInfo.html>.
- [45] M. G. Marshall, R. G. Ted, and H. Barbara, *State Failure Task Force Problem Set*, 2002, <http://www.cidcm.umd.edu/inscr/st-fail/index.html>.
- [46] M. Alvarez, A. C. Jose, L. Fernando, and P. Adam, *ACLP Political and Economic Database*, 1999, <http://www.ssc.upenn.edu/cheibub/data/>.
- [47] D. M. Titterton, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, Chichester, UK, 1985.
- [48] H.-T. Zhu and S.-Y. Lee, "A Bayesian analysis of finite mixtures in the LISREL model," *Psychometrika*, vol. 66, no. 1, pp. 133–152, 2001.
- [49] S.-Y. Lee and X.-Y. Song, "Maximum likelihood analysis of a general latent variable model with hierarchically mixed data," *Biometrics*, vol. 60, no. 3, pp. 624–636, 2004.
- [50] J. S. Liu, "The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem," *Journal of the American Statistical Association*, vol. 89, no. 427, pp. 958–966, 1994.
- [51] B. Nandram and M.-H. Chen, "Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence," *Journal of Statistical Computation & Simulation*, vol. 54, no. 1–3, pp. 129–144, 1996.
- [52] M. K. Cowles, "Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models," *Statistics and Computing*, vol. 6, no. 2, pp. 101–111, 1996.