Supplementary Material to "Bayesian Analysis of Two-part Nonlinear Latent Variable Model: Semiparametric Method"

Jian-Wei Gou

Department of Applied Mathematics, Nanjing Forestry University, Nanjing, China

Ye-Mao Xia

Department of Applied Mathematics, Nanjing Forestry University, Nanjing, China

De-Peng Jiang

Department of Community Health Sciences, University of Manitoba, Manitoba, Canada

S1. Bayesian semiparametric modeling for TNLVPM

S1.1 Truncated Dirichlet process prior for ϕ

Following the same notation in the main text and recalling that the truncated Dirichlet process (TDP) prior for $\mathcal{P}$ is given by

$$\mathcal{P}(\cdot) = \sum_{g=1}^{G} \pi_g \delta_{\boldsymbol{\phi}_g^*}(\cdot), \tag{S.1}$$

where $G \geq 1$ is any positive integer, $\boldsymbol{\phi}_g^* = \{\boldsymbol{\alpha}_{g1}^*, \boldsymbol{\alpha}_{g2}^*, \boldsymbol{\Psi}_{\epsilon g}^*, \boldsymbol{\mu}_g^*, \boldsymbol{\Psi}_{\delta g}^*\}$ are the iid. atoms with the common distribution H_0, and π_g are the random weights constructed via the stick-breaking procedure (Sethurmann, 1994)

$$\pi_1 = v_1, \quad \pi_g = v_g \prod_{\ell=1}^{g-1}(1 - v_\ell) \ \ (g = 2, \cdots, G-1), \quad \pi_G = \prod_{\ell=1}^{G-1}(1 - v_\ell), \tag{S.2}$$

in which $v_1, \cdots, v_{G-1}$ are the iid. $Beta(1, c)(c > 0)$ random variables. As an non-parametric Bayesian prior, TDP (S.1) was formally proposed by Ishwaran and Zarepour (2000) in approximating the well-known Dirichlet process prior (Ferguson, 1973).

Ishwaran and James (2001) further investigated its theoretical properties and exploited some applications in the context of univariate normal mixture model. The underlying advantages of using (S.1) are that it yields relatively simple Gibbs updates and allows one to draw values of $\mathcal{P}$ from the posterior directly.

Implementation of TDP requires specifying the truncated level G, the concentration parameter c and the baseline distribution H_0. In the context of normal mixture model, Ishwaran and James (2001) showed that even for large sample sizes, a mere truncation of $G = 50$ provides sufficient accuracy in approximating hierarchical model with $G = \infty$. In this paper, we take $G = 100$ in our empirical study. We also investigate the performance of other choices of G and find the results are not sensitive to these values. Moreover, as discussed in the main text, the values of c should be selected with care since they directly control the amount of clustering. We assign a gamma prior $Gamma(\nu_0, \lambda_0)$ to c with small ν_0 and λ_0. This routine favors both small and large values for c.

Based on the nature of problem, it is naturally to take H_0 as the form of

$$\mathrm{d}H_0 = p(\boldsymbol{\alpha}^*_{g1})p(\boldsymbol{\alpha}^*_{g2}, \boldsymbol{\Psi}^*_{\epsilon g})p(\boldsymbol{\mu}^*_g, \boldsymbol{\Psi}^*_{\delta g})\mathrm{d}\boldsymbol{\phi}^*_g, \tag{S.3}$$

in which

$$\begin{aligned}
&p(\boldsymbol{\alpha}^*_{g1}) \overset{D}{=} N_r(\boldsymbol{\alpha}_{10}, \mathbf{A}_{10}),\\
&p(\boldsymbol{\alpha}^*_{g2}, \boldsymbol{\Psi}^*_{\epsilon g}) = p(\boldsymbol{\Psi}^*_{\epsilon g})p(\boldsymbol{\alpha}^*_{g2}|\boldsymbol{\Psi}^*_{\epsilon g}) \overset{D}{=} \prod_{j=1}^{r} IG(\alpha_{\epsilon 0j}, \beta_{\epsilon 0j}) \times N(\alpha_{20j}, \psi^*_{\epsilon gj}A_{20j}),\\
&p(\boldsymbol{\mu}^*_g, \boldsymbol{\Psi}^*_{g\delta}) = p(\boldsymbol{\Psi}^*_{\delta g})p(\boldsymbol{\mu}^*_g|\boldsymbol{\Psi}^*_{\delta g}) \overset{D}{=} \prod_{j=1}^{s} IG(\alpha_{\delta j0}, \beta_{\delta j0}) \times N(\mu_{j0}, \psi^*_{\delta gj}A_{\mu 0j}),
\end{aligned} \tag{S.4}$$

where $IG(a, b)$ is the inverse Gamma distribution with scale $a > 0$ and shape $b > 0$; The hyper-parameters $\boldsymbol{\alpha}_{10}$, $\mathbf{A}_{10}$, $\alpha_{\epsilon 0j}$, $\beta_{\epsilon 0j}$, α_{20j}, A_{20j}, $\alpha_{\delta j0}$, $\beta_{\delta j0}$, μ_{j0} and $A_{\mu 0j}$ are treated as fixed and known. In practice, the values of hyper-parameters are often taken to ensure that the distributions in (S.4) behave as the non-informative priors. It is worth noting that unlike that in the parametric Bayesian analysis, a degenerated or non-informative distribution for H_0 is prohibitive.

S1.2 Polya-Gamma stochastic expression and re-parametrization

We develop a Bayesian procedure for analyzing the proposed model. The MCMC sampling method is used to conduct posterior analysis. To facilitate the posterior sampling, we follow the same idea in Polson, Scott and Windle (2013) and rewrite the logistic model in Part one as follows

$$\frac{e^{u_{ij}\eta_{cij}}}{1+e^{\eta_{cij}}} = 2^{-1}\exp\{\kappa_{ij}\eta_{cij}\}\int_0^\infty \exp\left\{-\frac{1}{2}u_{ij}^*\eta_{cij}^2\right\} p_{\mathrm{PG}}(u_{ij}^*)\mathrm{d}u_{ij}^*, \qquad \text{(S.5)}$$

where $\kappa_{ij} = u_{ij} - 1/2$ and $\eta_{cij} = \alpha_{1ij} + \mathbf{x}_{1i}^T\boldsymbol{\gamma}_{1j} + \boldsymbol{\beta}_{1j}^T\boldsymbol{\omega}_i$; $p_{\mathrm{PG}}(\cdot)$ is the probability density function of standard Polya-Gamma distribution PG(1,0)(Polson, Scott and Windle, 2013), specified via

$$\mathrm{PG}(1,0) \overset{D}{=} \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{1}{(k-1/2)^2}E_k,$$

where E_ks are the iid. standard exponential random variables; 'D' denotes the identical distribution. In this case, the logistic model can be treated as the marginal density of the following joint distribution

$$p(u_{ij}, u_{ij}^*|\boldsymbol{\omega}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}) = 2^{-1}\exp\left\{\kappa_{ij}\eta_{cij} - \frac{1}{2}u_{ij}^*\eta_{cij}^2\right\} p_{\mathrm{PG}}(u_{ij}^*).$$

Note that $p_{\mathrm{PG}}(u_{ij}^*)$ does not involve $\boldsymbol{\omega}_i$, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}_i$, hence, the conditional distribution of $\boldsymbol{\omega}_i$, given u_{ij}, u_{ij}^*, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}_i$, is independent of $p_{\mathrm{PG}}(u_{ij}^*)$ and has a closed form. This will facilitate the routine coding and avoid tediously tuning in the Metropolis-Hastings(MH, Metropolis et al., 1953; Hastings, 1970) algorithm. However, a disadvantage in the expression (S.5) is that the full conditional $p(u_{ij}^*|u_{ij}, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\phi}_i)$ is not standard since it involves the sum of infinite many terms. This invokes the difficulty in simulating observations. The problem can be addressed via indirect sampling methods such as the rejection-acceptance sampling (Gilks and Wild, 1992; Polson, Scott and Windle, 2013) or the slice sampling algorithm (Walker, 2007) .

The semiparametric model (S.2) can be re-parameterized further. Let $K_i (i = 1, \cdots, N)$ be the indicator variable which takes value in $\{1, \cdots, G\}$ such that $\boldsymbol{\phi}_i =$

$\boldsymbol{\phi}^*_{K_i}$. Then, we can rewrite our formulation as the following hierarchy

$$
\begin{aligned}
&p(u_{ij}, u^*_{ij}|\boldsymbol{\Xi}^*, K_i, \boldsymbol{\omega}_i) \overset{D}{=} 2^{-1}\exp\left\{\kappa_{ij}\eta_{cij} - \frac{1}{2}u^*_{ij}\eta^2_{cij}\right\} p_{\mathrm{PG}}(u^*_{ij}),\\
&p(z_{ij}|\boldsymbol{\Xi}^*, K_i, u_{ij}=1, \boldsymbol{\omega}_i) \overset{D}{=} N(\eta_{zij}, \psi_{\epsilon ij}),\\
&p(\mathbf{v}_i|\boldsymbol{\Xi}^*, K_i, \boldsymbol{\omega}_i) \overset{D}{=} N_s(\boldsymbol{\mu}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i, \boldsymbol{\Psi}_{\delta i}),\\
&p(\boldsymbol{\eta}_i|\boldsymbol{\xi}_i) \overset{D}{=} N_{m_1}(\mathbf{B}_0^{-1}\boldsymbol{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i), \mathbf{B}_0^{-1}\boldsymbol{\Psi}_\delta^{-1}\mathbf{B}_0^{-T}),\\
&p(\boldsymbol{\xi}_i|\boldsymbol{\Phi}) \overset{D}{=} N_{m_2}(\mathbf{0}, \boldsymbol{\Phi}),\\
&K_i = \cdot|\boldsymbol{\pi} \overset{iid.}{\sim} \sum_{g=1}^{G}\pi_g\delta_g(\cdot),\\
&\boldsymbol{\phi}^*_g, iid. \sim H_0, \quad (\boldsymbol{\pi}|c) \sim p(\boldsymbol{\pi}|c),
\end{aligned}
\tag{S.6}
$$

where $\boldsymbol{\Xi}^* = \{\boldsymbol{\phi}^*_1, \cdots, \boldsymbol{\phi}^*_G\}$ is the collection of atoms, $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_G)$ is the vector of random weights and $\eta_{zij} = \alpha_{2ij} + \mathbf{x}^T_{2i}\boldsymbol{\gamma}_{2j} + \boldsymbol{\beta}^T_{2j}\boldsymbol{\omega}_i$.

Let $\mathbf{U} = \{u_{ij}\}$, $\mathbf{Z} = \{z_{ij}\}$, $\mathbf{V} = \{\mathbf{v}_i\}$ be the sets of observed variables, and $\mathbf{U}^* = \{u^*_{ij}\}_{i=1}^N$, $\boldsymbol{\Omega} = \{\boldsymbol{\omega}_i\}_{i=1}^N$ and $\mathbf{K} = \{K_1, \cdots, K_N\}$ be the collections of latent variables. It can be shown that conditional on c and $\boldsymbol{\theta}$, the joint distribution of $\{\mathbf{U}^*, \boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\pi}, \boldsymbol{\Xi}^*\}$ is given by

$$
\begin{aligned}
&p(\mathbf{U}, \mathbf{Z}, \mathbf{V}, \mathbf{U}^*, \boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\pi}, \boldsymbol{\Xi}^*|\boldsymbol{\theta}, c) \propto p(\mathbf{U}, \mathbf{U}^*|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta})\\
&\times p(\mathbf{Z}|\mathbf{U}, \boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) p(\mathbf{V}|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) p(\boldsymbol{\Omega}|\boldsymbol{\theta}) p(\mathbf{L}|\boldsymbol{\pi}) p(\boldsymbol{\pi}|c) p(\boldsymbol{\Xi}^*),
\end{aligned}
$$

where the fixed covariates are suppressed for notation compactness. The observed likelihood can be obtained by integrating out the latent quantities, which yields the complicated form due to high dimensional integrals present.

S1.3 Model identification

In this section, we will show that the proposal models involved in the simulation study and real example are identified with their free parameters $\boldsymbol{\theta}$. The method relies on the well-known *two-indicator rule* in the CFA model (see, e.g., Bollen, 1989, pp.247) and the stochastic representation of the logistic regression model (Holmes and Held, 2006).

We re-express the logistic regression model (Part one) as the hierarchical form:

$$\begin{aligned} u_i &= \begin{cases} 1 & u_j^* > 0 \\ 0 & u_i^* \leq 0 \end{cases}, \\ u_i^* &= \alpha_1 + \mathbf{x}_{1i}^T \boldsymbol{\gamma}_1 + \boldsymbol{\beta}_1^T \boldsymbol{\omega}_i + \epsilon_i^*, \\ \epsilon_i^*, &iid. \sim Lo(1), \end{aligned} \tag{S.7}$$

where $Lo(1)$ is the standard logistic distribution and ϵ_i^* is independent of latent factors and other error variables.

Let $\mathbf{w}_i^* = (u_i^*, z_i, \mathbf{v}_i^T)^T$, $\boldsymbol{\varepsilon}_i^* = (\epsilon_i^*, \epsilon_i, \boldsymbol{\delta}_i^T)^T$, and write

$$\boldsymbol{\alpha}_{all} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \boldsymbol{\mu} \end{pmatrix}, \ \boldsymbol{\gamma}_{all} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \mathbf{0} \end{pmatrix}, \boldsymbol{\Lambda}_{all} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\Lambda} \end{pmatrix}, \text{ and } \boldsymbol{\Psi}_{all} = \begin{pmatrix} \pi^2/3 & 0 & \mathbf{0} \\ 0 & \psi_\epsilon & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_\delta \end{pmatrix}.$$

The joint distribution of $\mathbf{w}_i^*$ is given by

$$\mathbf{w}_i^* = \boldsymbol{\alpha}_{all} + \boldsymbol{\gamma}_{all}\mathbf{x}_i + \boldsymbol{\Lambda}_{all}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i^*. \tag{S.8}$$

Here, $\mathbf{x}_{1i} = \mathbf{x}_{2i}$ by the simulation design and the real example. The first two moments of $\mathbf{w}_i$ are given by

$$E\mathbf{w}_i^* = \boldsymbol{\alpha}_{all} + \boldsymbol{\gamma}_{all}\mathbf{x}_i + \boldsymbol{\Lambda}_{all}\boldsymbol{\mu}_\omega, \quad \text{Cov}(\mathbf{w}_i^*) = \boldsymbol{\Lambda}_{all}\boldsymbol{\Sigma}_\omega\boldsymbol{\Lambda}_{all}^T + \boldsymbol{\Psi}_{all}, \tag{S.9}$$

where $\boldsymbol{\mu}_\omega$ and $\boldsymbol{\Sigma}_\omega$ are the mean and covariance matrix of $\boldsymbol{\omega}_i$ respectively.

Note that the observed likelihood $p(u_i, z_i, \mathbf{v}_i|\boldsymbol{\theta})$ is uniquely determined by $p(\mathbf{w}_i^*|\boldsymbol{\theta})$, hence, we are only required to show that $\boldsymbol{\theta}$ is uniquely determined based on $p(\mathbf{w}_i^*|\boldsymbol{\theta})$.

We first show that the structural parameters $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}_\zeta$ involved in the nonlinear structure equation are identified assuming that the mean and covariance matrix of $\boldsymbol{\omega}_i$ are uniquely determined by the model. As the matter of fact, recalling that $\mathbf{F}(\boldsymbol{\xi}_i) = (\xi_{i1}, \xi_{i2}, \xi_{i1}\xi_{i2})^T$ with $(\xi_{i1}, \xi_{i2}) \sim N_2(\mathbf{0}, \boldsymbol{\Phi})$ and $\zeta_i \sim N(0, \psi_\zeta)$, it can be shown that $\boldsymbol{\mu}_\omega = (\Gamma_3\phi_{12}, \mathbf{0}_2^T)^T$, and

$$\boldsymbol{\Sigma}_\omega = \begin{pmatrix} \tau + \psi_\zeta & \Gamma_1\phi_{11} + \Gamma_2\phi_{12} & \Gamma_1\phi_{12} + \Gamma_2\phi_{22} \\ & \phi_{11} & \phi_{12} \\ sym. & & \phi_{22} \end{pmatrix}, \tag{S.10}$$

where $\tau = \Gamma_1^2\phi_{11} + \Gamma_2^2\phi_{22} + \Gamma_3^2 E\xi_{i1}^2\xi_{i2}^2 + 2\Gamma_1\Gamma_2\phi_{12}$; $E\xi_{i1}^2\xi_{i2}^2$ is the second-order mixed moment of ξ_{i1} and ξ_{i2}, only depending on the elements in $\boldsymbol{\Phi}$. Obviously, if $\boldsymbol{\mu}_\omega$ and $\boldsymbol{\Sigma}_\omega$ are identified, then the structure parameter $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}_\zeta$ are all identified.

Next, we show that the unknown parameters including $\boldsymbol{\mu}_\omega$ and $\boldsymbol{\Sigma}_\omega$ involved in (S.8) are identified.

(i) For the simulation study, note that $\boldsymbol{\Lambda}$ is orthogonal in columns with $\Lambda_{11} = \Lambda_{32} = \Lambda_{52} = 1$ and $\boldsymbol{\Psi}_\delta$ is set to be diagonal. It follows from the *two-indicator rule* in the CFA that $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}_\delta$ and $\boldsymbol{\Sigma}_\omega$ are identified. This induces the identification of $\boldsymbol{\beta}_j (j = 1, 2)$ by the identification of covariance of $\mathbf{w}_i^*$. Moreover, note that $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_N)^T$ is full of rank in column according to the simulation design, the identification of $\boldsymbol{\alpha}_j$, $\boldsymbol{\gamma}_j$ and $\boldsymbol{\mu}_\omega$ are obtained immediately.

(ii) For the real example, note that

$$\boldsymbol{\Lambda}_{all} = \begin{pmatrix} 1 & \beta_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Lambda_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Lambda_{43} & \Lambda_{53} \end{pmatrix} \tag{S.11}$$

and $\boldsymbol{\Psi}_{all}$ is diagonal. The identification of $\boldsymbol{\Lambda}_{all}$, $\boldsymbol{\Psi}_{all}$ and $\boldsymbol{\Sigma}_\omega$ follows still from the *two-indicator rule*, which indicates the determinacy of $\boldsymbol{\beta}_j$, $\boldsymbol{\Lambda}$, ψ_ϵ and $\boldsymbol{\Psi}_\delta$. The identification of α_j, $\boldsymbol{\gamma}_j$ and $\boldsymbol{\mu}_\omega$ are obtained similar to (i).

S1.4 Issue on the number of factors.

Another issue in the current analysis involves the determinacy of the number of factors. This problem is usually formulated within the framework of exploratory factor analysis (EFA) and the confirmatory factor analysis (CFA). The basic formulation of EFA is that for a given set of response variables, one wants to search a fewer number of uncorrelated latent factors that will account for the correlations of the response variables so that when the latent factors are partialled out from the response variables, there no longer remain any correlations between them. The number of factors is not preassigned and usually determined via the model selection procedure. The factor rotations are also required to obtain better interpretations. CFA is an extension of EFA and is developed by Jöreskog and Sörborm (1996). In this context, the experimenter has already obtained from the substantive theory or the exploratory analysis a certain amount of knowledge about the model and is in a position to formulate a more precise model. The factors and their number are determined in advance. The model is identified via fixing parameters, the resulting solution is

directly interpretable, and the subsequent rotation of factor loading matrix is not necessary. Goodness-of-fits of the posited models are compared to establish a plausible model for substantive theory in the real world. Our development is along the lines of the CFA and the number of factors is specified in advance based on the nature of the problem under consideration.

S2. Markov Chain Monte Carlo sampling and full Conditionals

S2.1 MCMC sampling scheme

The sampling scheme in the blocked Gibbs sampler includes four types of moves: updating the components involved in the two-part model, the components in the factor analysis model, the components in the structural equation and the components in the semiparametric model, which can be formulated as the following conditional distributions:

(i) $p(\mathbf{U}^*|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U})$,

(ii) $p(\boldsymbol{\Omega}|\mathbf{U}^*, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{V})$,

(iii) $p(\boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{U}^*, \mathbf{U}, \mathbf{Z}, \mathbf{V})$,

(iv) $p(\boldsymbol{\pi}, \boldsymbol{\Xi}^*|\mathbf{K}, \mathbf{U}^*, \boldsymbol{\theta}, \mathbf{Z}, \mathbf{V})$,

(v) $p(\mathbf{K}|\mathbf{U}^*, \boldsymbol{\pi}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{Z}, \mathbf{V})$, and

(vi) $p(c|\boldsymbol{\pi})$,

in which the redundant variables are removed from the conditioning set either by explicit integration or by conditional independence. Note that except for updating $\boldsymbol{\Omega}$, the conditionals involved in all moves are the standard distributions such as normal, Gamma and inverse-Gamma. Hence, simulating observations from these distributions are relatively easy and fast. However, drawing $\boldsymbol{\Omega}$ is not straightforward since the nonlinear function $\mathbf{F}(\cdot)$ are presented. We implement MH algorithm to tackle this problem.

S2.2 Full conditionals

In this section, we will present the technical details on the full conditional distributions involved in the blocked Gibbs sampler.

(i) $p(\mathbf{U}^*|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U})$;

As discussed in S1.2, conditioning upon $\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}$ and $\mathbf{U}$, the posterior distribution of $\mathbf{U}^*$ does not depend on $\mathbf{Z}$ and $\mathbf{V}$, and is still the Polya-Gamma distribution. By some algebra calculations, we have

$$p(\mathbf{U}^*|\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{U}) = \prod_{i=1}^{N}\prod_{j=1}^{r} p(u_{ij}^*|\boldsymbol{\theta}, \boldsymbol{\omega}_i), \text{ and}$$
$$p(u_{ij}^*|\boldsymbol{\theta}, \boldsymbol{\omega}_i) \stackrel{D}{=} \mathrm{PG}(1, \alpha_{1ij} + \widetilde{\boldsymbol{\gamma}}_{1j}^T \boldsymbol{\omega}_{1i}^*). \tag{S.12}$$

By taking advantage of rejection sampling method, Polson, Scott and Windle (2013) devised an efficient algorithm for generating observations from this target distribution, see Polson, Scott and Windle (2013) for more details. Their algorithm is adopted here to draw $\mathbf{U}^*$.

(ii) $p(\boldsymbol{\Omega}|\mathbf{Y}^*, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{Z}, \mathbf{U}, \mathbf{V})$.

Based on (S.6), it can be shown that

$$p(\boldsymbol{\Omega}|\mathbf{U}^*, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{Y}) = \prod_{i=1}^{N} p(\boldsymbol{\omega}_i|\mathbf{U}^*, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{Y}), \tag{S.13}$$

where $p(\boldsymbol{\omega}_i|\mathbf{U}^*, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{Y}) \propto p(\mathbf{u}_i, \mathbf{u}_i^*|\boldsymbol{\omega}_i, K_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) p(\mathbf{z}_i|\mathbf{u}_i, \boldsymbol{\omega}_i, K_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) p(\mathbf{v}_i|\boldsymbol{\omega}_i, K_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i|\boldsymbol{\xi}_i, \boldsymbol{\theta}) p(\boldsymbol{\xi}_i|\boldsymbol{\Phi})$. Hence, we can generate $\boldsymbol{\Omega}$ by drawing $\boldsymbol{\omega}_i$ independently from $p(\boldsymbol{\omega}_i|\mathbf{Y}^*, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{Y})$. However, due to that nonlinear function $\mathbf{F}(\cdot)$ in the structural equation is involved, no closed form can be available for this target distribution. We implement the Metropolis and Hastings algorithm. Specifically, given the current value $\boldsymbol{\omega}_i^{(h)}$ at the hth iteration in the MH algorithm, we draw a candidate $\boldsymbol{\omega}_i^*$ from the proposal distribution $N_m(\boldsymbol{\omega}_i^{(h)}, \tau\boldsymbol{\Sigma}_i^*)$, where $\boldsymbol{\Sigma}_i^* = (\boldsymbol{\beta}_1^T\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2^T\boldsymbol{\Psi}_{\epsilon i}^{*-1}\boldsymbol{\beta}_2 + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}_{\delta i}^{-1}\boldsymbol{\Lambda} + \boldsymbol{\Sigma}_\omega^{-1})^{-1}$, and

$$\boldsymbol{\Sigma}_\omega^{-1} = \begin{pmatrix} \mathbf{B}_0^{-T}\boldsymbol{\Psi}_\zeta^{-1}\mathbf{B}_0^{-1} & -\mathbf{B}_0^{-T}\boldsymbol{\Psi}_\zeta^{-1}\boldsymbol{\Gamma}\boldsymbol{\Delta} \\ -\boldsymbol{\Delta}^T\boldsymbol{\Gamma}^T\boldsymbol{\Psi}_\zeta^{-1}\mathbf{B}_0^{-1} & \boldsymbol{\Phi}^{-1} + \boldsymbol{\Delta}^T\boldsymbol{\Gamma}^T\boldsymbol{\Psi}_\zeta^{-1}\boldsymbol{\Gamma}\boldsymbol{\Delta} \end{pmatrix}$$

with $\boldsymbol{\Delta} = \partial\mathbf{F}(\boldsymbol{\xi}_i)/\partial\boldsymbol{\xi}_i^T\,|_{\boldsymbol{\xi}_i=\mathbf{0}}$ and $\boldsymbol{\Psi}_{\epsilon i}^{*-1} = diag\{u_{i1}/\psi_{\epsilon i1}, \cdots, u_{ir}/\psi_{\epsilon ir}\}$, then we accept this candidate $\boldsymbol{\omega}_i^*$ as $\boldsymbol{\omega}_i^{(h+1)}$ with the probability

$$\min\left\{1, \frac{p(\boldsymbol{\omega}_i^*|\mathbf{U}^*, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{Y})}{p(\boldsymbol{\omega}_i^{(h)}|\mathbf{U}^*, \mathbf{K}, \boldsymbol{\Xi}^*, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}, \mathbf{Y})}\right\}.$$

The quantity τ is chosen so that the average rate is approximately 0.45 (see Cowles, 1996).

(iii) $p(\boldsymbol{\theta}|\mathbf{U}^*, \boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{U}, \mathbf{Z}, \mathbf{V})$

Note that $\boldsymbol{\theta} = \{\widetilde{\boldsymbol{\gamma}}_1, \widetilde{\boldsymbol{\gamma}}_2, \boldsymbol{\Lambda}, \boldsymbol{\Pi}, \boldsymbol{\Psi}_\delta, \boldsymbol{\Phi}\}$. Hence, we can achieve the draw of $\boldsymbol{\theta}$ by drawing sequently (1) $\widetilde{\boldsymbol{\gamma}}_1$ from $p(\widetilde{\boldsymbol{\gamma}}_1|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{U}^*, \mathbf{U})$; (2) $\widetilde{\boldsymbol{\gamma}}_2$ from $p(\widetilde{\boldsymbol{\gamma}}_2|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{U}, \mathbf{Z})$; (3) $\boldsymbol{\Lambda}$ from $p(\boldsymbol{\Lambda}|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{V})$; (4) $\{\boldsymbol{\Pi}, \boldsymbol{\Psi}_\delta\}$ from $p(\boldsymbol{\Pi}, \boldsymbol{\Psi}_\delta|\boldsymbol{\Omega})$ and (5) $\boldsymbol{\Phi}$ from $p(\boldsymbol{\Phi}|\boldsymbol{\Omega})$.

Firstly, let $\mathbf{x}_{1i}^* = (\mathbf{x}_{1i}^T, \boldsymbol{\omega}_i^T)^T$. A direct calculation shows that

$$p(\widetilde{\boldsymbol{\gamma}}_1|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{U}^*, \mathbf{U}) = \prod_{j=1}^{r} p(\widetilde{\boldsymbol{\gamma}}_{1j}|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{U}^*, \mathbf{U}) \stackrel{D}{=} \prod_{j=1}^{r} N_{q_1+m}(\widehat{\boldsymbol{\mu}}_{\gamma_1 j}, \widehat{\boldsymbol{\Sigma}}_{\gamma_1 j}), \quad \text{(S.14)}$$

in which

$$\widehat{\boldsymbol{\mu}}_{\gamma_1 j} = \widehat{\boldsymbol{\Sigma}}_{\gamma_1}^{-1}\left(\widetilde{\boldsymbol{\Sigma}}_{1j0}^{-1}\widetilde{\boldsymbol{\gamma}}_{1j0} + \sum_{i=1}^{N} \mathbf{x}_{1i}^*(\kappa_{ij} - u_{ij}^*\alpha_{1ij})\right), \quad \widehat{\boldsymbol{\Sigma}}_{\alpha_1}^{-1} = \widetilde{\boldsymbol{\Sigma}}_{1j0}^{-1} + \sum_{i=1}^{N} \mathbf{x}_{1i}^*\mathbf{x}_{1i}^{*T}.$$

Secondly, let $\mathbf{x}_{2i}^* = (\mathbf{x}_{2i}^T, \boldsymbol{\omega}_i^T)^T$. Similar to update $\widetilde{\boldsymbol{\gamma}}_1$, we can follow the standard regression analysis procedure and show that

$$p(\widetilde{\boldsymbol{\gamma}}_2|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{U}, \mathbf{Z}) \stackrel{D}{=} \prod_{j=1}^{r} N_{q_2+m}(\widehat{\boldsymbol{\mu}}_{\gamma 2 j}, \widehat{\boldsymbol{\Sigma}}_{\gamma 2 j}), \quad \text{(S.15)}$$

where

$$\widehat{\boldsymbol{\mu}}_{\gamma 2 j} = \widehat{\boldsymbol{\Sigma}}_{\gamma 2 j}\left(\widetilde{\boldsymbol{\Sigma}}_{2j0}^{-1}\widetilde{\boldsymbol{\gamma}}_{2j0} + \sum_{i=1}^{N} \mathbf{x}_{2i}^* u_{ij}(z_{ij} - \alpha_{2ij})\right), \ \widehat{\boldsymbol{\Sigma}}_{\gamma 2 j}^{-1} = \widetilde{\boldsymbol{\Sigma}}_{2j0}^{-1} + \sum_{i=1}^{N} u_{ij}\mathbf{x}_{2i}^*\mathbf{x}_{2i}^{*T}.$$

Thirdly, for ease of exposition, we introduce the following notation

$$\mathbf{S}_{\omega v k} = \sum_{g=1}^{G} \boldsymbol{\omega}_i(v_{ik} - \mu_{ik}), \quad \mathbf{S}_{\omega\omega} = \sum_{g=1}^{G} \boldsymbol{\omega}_i\boldsymbol{\omega}_i^T, \text{ and } S_{vvk} = \sum_{g=1}^{G}(v_{gk} - \mu_k)^2.$$

By some algebra, it can be shown that

$$p(\boldsymbol{\Lambda}|\boldsymbol{\Omega}, \mathbf{K}, \boldsymbol{\Xi}^*, \mathbf{V}) = \prod_{k=1}^{s} p(\boldsymbol{\Lambda}_k|\psi_{\delta k}, \boldsymbol{\Omega}, \mathbf{V}) \stackrel{D}{=} \prod_{k=1}^{s} N_m(\widehat{\mathbf{a}}_{\delta k}, \widehat{\mathbf{H}}_{\delta k}), \quad \text{(S.16)}$$

where

$$\widehat{\mathbf{a}}_{\delta k} = \widehat{\mathbf{H}}_{\delta k}[\mathbf{H}_{\delta 0 k}^{-1}\boldsymbol{\Lambda}_{0k} + \psi_{\delta k}\mathbf{S}_{\omega v k}], \quad \widehat{\mathbf{H}}_{\delta k} = (\mathbf{H}_{\delta 0 k}^{-1} + \psi_{\delta k}\mathbf{S}_{\omega\omega k})^{-1}.$$

Fourthly, let $\mathbf{G}_i = \mathbf{G}(\boldsymbol{\omega}_i)$ and

$$\mathbf{S}_{G\eta\ell} = \sum_{i=1}^{N} \mathbf{G}_i \eta_{i\ell}, \quad \mathbf{S}_{GG} = \sum_{i=1}^{N} \mathbf{G}_i \mathbf{G}_i^T, \text{ and } S_{\eta\eta\ell} = \sum_{i=1}^{N} \eta_{i\ell}^2.$$

Similar to that in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}_\delta$, it can be shown that

$$\begin{aligned} p(\boldsymbol{\Pi}, \boldsymbol{\Psi}_\zeta | \boldsymbol{\Omega}) &= \prod_{\ell=1}^{m_1} p(\psi_{\zeta\ell}|\boldsymbol{\Omega}) p(\boldsymbol{\Pi}_\ell | \psi_{\zeta\ell}, \boldsymbol{\Omega}) \\ &\stackrel{D}{=} \prod_{\ell=1}^{m_1} IG(\widehat{\alpha}_{\zeta\ell}, \widehat{\beta}_{\zeta\ell}) \times N_m(\widehat{\mathbf{a}}_{\zeta\ell}, \psi_{\zeta\ell}\widehat{\mathbf{H}}_{\zeta\ell}), \end{aligned} \tag{S.17}$$

where

$$\begin{aligned} &\widehat{\mathbf{a}}_{\zeta\ell} = \widehat{\mathbf{H}}_{\zeta\ell}[\mathbf{H}_{\zeta 0\ell}^{-1}\boldsymbol{\Pi}_{0\ell} + \mathbf{S}_{G\eta k}], \quad \widehat{\mathbf{H}}_{\zeta\ell} = (\mathbf{H}_{\delta 0\ell}^{-1} + \mathbf{S}_{GG})^{-1}, \\ &\widehat{\alpha}_{\zeta\ell} = \alpha_{\zeta 0\ell} + N/2, \\ &\widehat{\beta}_{\zeta\ell} = \beta_{\zeta 0\ell} + \left\{\boldsymbol{\Pi}_{0\ell}^T \mathbf{H}_{\zeta 0\ell}^{-1} \boldsymbol{\Pi}_{0\ell} + S_{\eta\eta\ell} - \widehat{\mathbf{a}}_{\zeta\ell}^T \widehat{\mathbf{H}}_{\zeta\ell}^{-1} \widehat{\mathbf{a}}_{\zeta\ell}\right\}/2. \end{aligned}$$

Lastly, let $\mathbf{S}_{\xi\xi} = \sum_{i=1}^{N} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T$. It can be shown straightforwardly that

$$p(\boldsymbol{\Phi}|\boldsymbol{\Omega}) \stackrel{D}{=} IW_{m_2}(\rho_0 + N, \widehat{\mathbf{R}}^{-1}) \tag{S.18}$$

where $\widehat{\mathbf{R}}^{-1} = \mathbf{R}_0^{-1} + \mathbf{S}_{\xi\xi}$.

(viii)$p(\boldsymbol{\pi}, \boldsymbol{\Xi}^* | \mathbf{K}, \boldsymbol{\Omega}, c, \mathbf{Y}^*, \mathbf{U}, \mathbf{Z}, \mathbf{V})$

To derive the conditional distributions corresponding to the mixing proportions $\boldsymbol{\pi}$ and atoms $\boldsymbol{\Xi}^*$, we notice that

$$p(\boldsymbol{\pi}, \boldsymbol{\Xi}^* | \mathbf{L}, \boldsymbol{\Omega}, c) = p(\boldsymbol{\pi}|\mathbf{K}, c) p(\boldsymbol{\Xi}^* | \mathbf{K}, \boldsymbol{\Omega}),$$

and $p(\boldsymbol{\pi}|\mathbf{K}, c) \propto p(\mathbf{K}|\boldsymbol{\pi}) p(\boldsymbol{\pi}|c)$. Hence drawing $\{\boldsymbol{\pi}, \boldsymbol{\Xi}^*\}$ can be accomplished by first drawing $\boldsymbol{\pi}$ from $p(\boldsymbol{\pi}|\mathbf{K}, c)$ and then drawing $\boldsymbol{\Xi}^*$ from $p(\boldsymbol{\Xi}^*|\mathbf{K}, \boldsymbol{\Omega})$. Let m_g denote the number of L_i such that $L_i = g$. By some algebraic calculations, it can be found that the full conditional distribution of $\boldsymbol{\pi}$ is given by

$$\pi_1 = v_1^*, \ \pi_g = \prod_{l<g}(1 - v_l^*), \ \pi_G = \prod_{l<G}(1 - v_l^*), \tag{S.19}$$

in which v_g^* are independently distributed with $Beta(1+m_g, c+\sum_{\ell=g+1}^{G} m_g)$. As a result, draw of $\boldsymbol{\pi}$ can be achieved by first sampling $G-1$ independent v_g^* and then transforming them via (S.19).

Let $\mathbf{K}^* = \{K_1^*, \cdots, K_m^*\}$ denote the set of unique values among $\mathbf{K}$ and $\boldsymbol{\Xi}^*_{\mathbf{K}^*} = \{\boldsymbol{\phi}^*_{K_1^*}, \cdots, \boldsymbol{\phi}^*_{K_m^*}\}$ be the distinct atoms; $\boldsymbol{\Xi}^*_{-\mathbf{K}^*}$ is the set of $\boldsymbol{\Xi}^*$ with $\boldsymbol{\Xi}^*_{-\mathbf{K}^*}$ removed. It is obvious that

$$p(\boldsymbol{\Xi}^*|\mathbf{K}, \boldsymbol{\Omega}) = p(\boldsymbol{\Xi}^*_{\mathbf{K}^*}|\mathbf{K}, \boldsymbol{\Omega}) p(\boldsymbol{\Xi}^*_{-\mathbf{K}^*}|\mathbf{K}, \boldsymbol{\Omega}). \tag{S.20}$$

Note that the components in $\boldsymbol{\Xi}^*_{-\mathbf{K}^*}$ are iid. with the baseline distribution H_0 and hence sampling is straightforward. Further, let m_g be the size of set $\{i : K_i = g\}$ and n^*_{gj} be the number of individuals such that $K_i = g$ and $u_{ij} = 1$. To derive the conditional distribution, we let

$$\begin{aligned}
&\bar{\boldsymbol{\kappa}}^*_g : \bar{\kappa}^*_{gj} = \sum_{i:K_i=g} (\kappa_{ij} - u^*_{ij}\widetilde{\boldsymbol{\gamma}}_1 \mathbf{x}^*_{1i}), \quad \bar{z}^*_{gj} = \sum_{i:K_i=g, u_{ij}=1} (z_{ij} - \widetilde{\boldsymbol{\gamma}}^T_{2j} \mathbf{x}^*_{2i}), \\
&S_{zzg} = \sum_{\{i:K_i=g, u_{ij}=1\}} (z_{ij} - \widehat{\boldsymbol{\gamma}}^T_j \mathbf{x}^*_{1i})^2, \quad S_{vvg} = \sum_{\{i:K_i=g, u_{ij}=1\}} (v_{ij} - \boldsymbol{\Lambda}^T_j \boldsymbol{\omega}_i)^2.
\end{aligned}$$

It can be shown that for any $g \in \mathbf{K}^*$,

$$\begin{aligned}
p(\boldsymbol{\alpha}^*_{1g}|\mathbf{K}, \boldsymbol{\Omega}, \mathbf{U}^*, \boldsymbol{\theta}, \mathbf{U}) &\overset{D}{=} N_r(\widehat{\mathbf{m}}_{\alpha_1 g}, \widehat{\boldsymbol{\Sigma}}_{\alpha_1 g}), \\
p(\boldsymbol{\alpha}^*_{2g}, \boldsymbol{\Psi}^*_{\epsilon g}|\mathbf{K}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}) &= \prod_{j=1}^{r} p(\alpha^*_{2gj}, \psi^*_{\epsilon gj}|\mathbf{K}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}) \\
&\overset{D}{=} \prod_{j=1}^{r} IG(\widehat{\alpha}_{\epsilon gj}, \widehat{\beta}_{\epsilon gj}) \times N(\widehat{m}_{\alpha_2 gj}, \psi^*_{\epsilon gj}\widehat{\sigma}^2_{\alpha_2 gj}), \\
p(\boldsymbol{\mu}_g, \boldsymbol{\Psi}^*_{\delta g}|\mathbf{K}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{V}) &= \prod_{k=1}^{s} p(\mu^*_{gk}, \psi^*_{\delta gk}|\mathbf{K}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{U}, \mathbf{Z}) \\
&\overset{D}{=} \prod_{k=1}^{s} IG(\widehat{\alpha}_{\delta k}, \widehat{\beta}_{\delta k}) \times N(\widehat{m}_{\mu k}, \psi^*_{\delta gk}\widehat{\sigma}^2_{\mu gk}),
\end{aligned}$$

in which

$$\widehat{\boldsymbol{\Sigma}}_{\alpha_1 g}^{-1} = (m_g \mathbf{I}_r + \mathbf{A}_{10}^{-1})^{-1}, \quad \widehat{\mathbf{m}}_{\alpha_1 g} = \widehat{\boldsymbol{\Sigma}}_{\alpha_1 g}\big[\mathbf{A}_{10}^{-1}\boldsymbol{\alpha}_{10} + m_g \bar{\boldsymbol{\kappa}}_g^*\big],$$
$$\widehat{\sigma}_{\alpha_2 gj}^2 = (n_{gj}^* + A_{20j}^{-1})^{-1}, \quad \widehat{m}_{\alpha_2 gj} = \widehat{\sigma}_{\alpha_2 gj}^2\big[A_{20j}^{-1}\alpha_{20j} + n_{gj}\bar{z}_{gj}^*\big],$$
$$\widehat{\alpha}_{\epsilon gj} = \alpha_{\epsilon 0j} + n_{gj}^*/2,$$
$$\widehat{\beta}_{\epsilon gj} = \beta_{\epsilon 0j} + (A_{20j}^{-1}\alpha_{20j}^2 + S_{zzg} - \widehat{m}_{\alpha_2 gj}^2/\widehat{\sigma}_{\alpha_2 gj}^2)/2;$$
$$\widehat{\sigma}_{\mu gk}^2 = (m_g^* + A_{\mu 0k}^{-1})^{-1}, \quad \widehat{m}_{\mu gk} = \widehat{\sigma}_{\mu gk}^2\big[A_{\mu 0k}^{-1}\mu_{0k} + n_{gj}\bar{v}_{gk}^*\big],$$
$$\widehat{\alpha}_{\delta gk} = \alpha_{\delta 0j} + n_{gk}^*/2,$$
$$\widehat{\beta}_{\delta gk} = \beta_{\delta 0j} + (A_{\mu 0k}^{-1}\mu_{0j}^2 + S_{vvk} - \widehat{m}_{\mu gk}^2/\widehat{\sigma}_{\mu gk}^2)/2.$$

(ix) $p(\mathbf{K}|\boldsymbol{\Omega}, \boldsymbol{\pi})$

Note that

$$p(\mathbf{K}|\mathbf{U}^*, \boldsymbol{\pi}, \boldsymbol{\Omega}, \mathbf{U}, \mathbf{Z}, \mathbf{V}, \boldsymbol{\theta}) = \prod_{i=1}^{N} p(K_i|\mathbf{U}^*, \boldsymbol{\pi}, \boldsymbol{\Omega}, \mathbf{U}, \mathbf{Z}, \mathbf{V}, \boldsymbol{\theta})$$

and

$$p(K_i|\mathbf{U}^*, \boldsymbol{\pi}, \boldsymbol{\Omega}, \mathbf{U}, \mathbf{Z}, \mathbf{V}, \boldsymbol{\theta}) \propto$$
$$p(\mathbf{u}_i^*, \mathbf{u}_i|\boldsymbol{\omega}_i, K_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) \prod_{j=1}^{r} p(z_{ij}|u_{ij} = 1, \boldsymbol{\omega}_i, K_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) p(\mathbf{v}_i|\boldsymbol{\omega}_i, K_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) p(K_i|\boldsymbol{\pi}),$$

hence, we have

$$(K_i = \cdot|\boldsymbol{\Omega}, \boldsymbol{\pi}) \overset{iid.}{\sim} \sum_{g=1}^{G} \pi_{ig}\delta_g(\cdot) \tag{S.21}$$

in which

$$\pi_{ig}^* = c_{ig}\pi_g p(\mathbf{u}_i^*, \mathbf{u}_i|\boldsymbol{\omega}_i, K_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) \prod_{j=1}^{r} p(z_{ij}|u_{ij} = 1, \boldsymbol{\omega}_i, K_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}) p(\mathbf{v}_i|\boldsymbol{\omega}_i, L_i, \boldsymbol{\Xi}^*, \boldsymbol{\theta}),$$

and c_{ig} is the normalization constant such that $\sum_{g=1}^{G} \pi_{ig}^* = 1.0$.

(x) $p(c|\boldsymbol{\pi})$

Similar to Xia and Gou (2016), it can be shown directly that

$$p(c|\boldsymbol{\pi}) \overset{D}{=} Ga\left(\nu_0 + G - 1, \lambda_0 - \sum_{g=1}^{G-1} \log(1 - v_g^*)\right) \tag{S.22}$$

in which v_g^*s are given in (S.19).

S3. Simulation study

In this section, simulation study is presented to assess the performance of the proposed procedure. The main objective is to assess the accuracy of estimates of unknown parameters and the adequacy of model fits when the posited models are correctly specified or deviated away from the true model. For this end, we take one semi-continuous variable ($r = 1$), six continuous manifest variables ($s = 6$) and three latent factors ($m = 3$), which are analogous to our real example. The nonlinear structural equation is taken as

$$\eta_i = \Gamma_1 \xi_{i1} + \Gamma_2 \xi_{i2} + \Gamma_3 \xi_{i1} \xi_{i2} + \zeta_i, \tag{S.23}$$

where $(\xi_{i1}, \xi_{i2}) \sim N_2(\mathbf{0}, \mathbf{\Phi})$ and $\zeta_i \sim N(0, \psi_\zeta)$. The fixed covariates involved in the model are taken to be identical: $\mathbf{x}_{1i} = \mathbf{x}_{2i} = (x_{1ij}, x_{2ij}, x_{3ij})^T$, in which x_{1ij}s are independently generated from uniform distribution on the unit interval $[0, 1]$, x_{2ij}s are independently drawn from standard normal distribution, whereas x_{3ij}s are independently generated from Bernoulli distribution with occurrence probability 0.3.

We considered three types of data which correspond to three different true population models. The first one is generated from single parameter model (denoted by 'PAR') given by

$$\begin{aligned} u_i &\sim p_{Lo}(u_i | \alpha_1 + \boldsymbol{\gamma}_1^T \mathbf{x}_i + \boldsymbol{\beta}_1^T \boldsymbol{\omega}_i), \\ z_i &\sim N(\alpha_2 + \boldsymbol{\gamma}_2^T \mathbf{x}_i + \boldsymbol{\beta}_2^T \boldsymbol{\omega}_i, \psi_\epsilon), \\ \mathbf{v}_i &\sim N_6(\boldsymbol{\mu} + \mathbf{\Lambda} \boldsymbol{\omega}_i, \mathbf{\Psi}_\delta), \end{aligned} \tag{S.24}$$

where $p_{Lo}(u|\vartheta)$ is the standard logistic density function with the predictor ϑ. The second data set (denoted by 'MIX' model) is generated from the two-component mixture model

$$\begin{aligned} u_i &\sim \sum_{k=1}^{2} \pi^{(k)} p_{Lo}(u_i | \alpha_1^{(k)} + \boldsymbol{\gamma}_1^T \mathbf{x}_i + \boldsymbol{\beta}_1^T \boldsymbol{\omega}_i), \\ z_i &\sim \sum_{k=1}^{2} \pi^{(k)} N(\alpha_2^{(k)} + \boldsymbol{\gamma}_2^T \mathbf{x}_i + \boldsymbol{\beta}_2^T \boldsymbol{\omega}_i, \psi_{\epsilon j}^{(k)}), \\ \mathbf{v}_i &\sim \sum_{k=1}^{2} \pi^{(k)} N_6(\boldsymbol{\mu}^{(k)} + \mathbf{\Lambda} \boldsymbol{\omega}_i, \mathbf{\Psi}_\delta^{(k)}), \end{aligned} \tag{S.25}$$

where the parameters with superscript '(k)' ($k = 1, 2$) are referred to as Component k; $\pi^{(k)}(k = 1, 2)$ are the mixing proportions satisfying: $\pi^{(k)} > 0$ and $\pi^{(1)} + \pi^{(2)} = 1$. The last one is drawn from the parametric model similar to 'PAR' but with heavy tails (denoted by 'HT') given by

$$\begin{aligned} u_i &\sim \mathcal{T}_2^+(u_i|\alpha_1 + \boldsymbol{\gamma}_1^T\mathbf{x}_i + \boldsymbol{\beta}_1^T\boldsymbol{\omega}_i), \\ z_i &\sim t_2(\alpha_1 + \boldsymbol{\gamma}_2^T\mathbf{x}_i + \boldsymbol{\beta}_2^T\boldsymbol{\omega}_i, \psi_\epsilon), \\ \mathbf{v}_i &\sim t_2(\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i, \boldsymbol{\Psi}_\delta), \end{aligned} \tag{S.26}$$

where t_ν denotes the t-distribution with degree of freedom ν and $\mathcal{T}_\nu^+$ is the indicator function of t_ν-distribution defined similar to the well-known Probit function, that is, $u_i = 1$ if $\mathcal{T}_2^+ > 0$ and 0 otherwise.

The true population values of unknown parameters in structural equation (S.23) are taken as $\boldsymbol{\Pi} = (0.0^*, 0.5, 0.5, 0.5)$, $\psi_\zeta = 1.0$ and

$$\boldsymbol{\Phi} = \begin{pmatrix} 1.0^* & 0.3 \\ 0.3 & 1.0^* \end{pmatrix}.$$

This produces that the true mean of $(\eta_i, \xi_{i1}, \xi_{i2})^T$ is $(0.3, 0.0, 0.0)^T$ and the covariance matrix

$$\begin{pmatrix} 1.9225 & & sym \\ 0.625 & 1.0 & \\ 0.625 & 0.3 & 1.0 \end{pmatrix}. \tag{S.27}$$

For the (S.24) and (S.26), the true population values of unknown parameters in Eq.s (S.24), (S.25) and (S.26) are taken as follows: $\alpha_1 = \alpha_2 = \mu_j = -1.5(j = 1, \cdots, 6)$, $\boldsymbol{\gamma}_1 = (0.7, 0.7, 0.7)^T$, $\boldsymbol{\gamma}_2 = (1.2, -0.7, 0.5)^T$, $\boldsymbol{\beta}_1 = (0.8, 0.8, 0.8)^T$, $\boldsymbol{\beta}_2 = (0.5, 0.5, 0.5)^T$, $\psi_\epsilon = 0.49$, $\boldsymbol{\Psi}_\delta = 0.49\mathbf{I}_6$ and

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1.0^* & 0.8 & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 1.0^* & 0.8 & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & 0.8 \end{pmatrix}.$$

where the elements with an asterisk are considered as fixed for model identification.

For 'MIX', the true values of component-invariant parameters are set as the same as those in (S.24) and (S.26), while the values of component-specific parameters are taken as in Component one in (S.25)while $\alpha_1^{(2)} = \alpha_2^{(2)} = \mu_j^{(2)} = -1.5$, $\alpha_1^{(2)} = \alpha_2^{(2)} =$

$\mu_j^{(2)} = 2.0(j = 1, \cdots, 6)$, $\psi_\epsilon^{(2)} = 0.49$, $\psi_\epsilon^{(2)} = 1.0$, $\mathbf{\Psi}_\delta^{(1)} = 0.49\mathbf{I}_6$ $\mathbf{\Psi}_\delta^{(2)} = \mathbf{I}_6$ and $(\pi^{(1)}, \pi^{(2)}) = (0.3, 0.7)$.

With the the settings given above, we generate the three data sets by first drawing latent factors from the structural equation (S.23) and then drawing continuous and semi-continuous observations from (S.24), (S.25) and (S.26) respectively. To investigate the effects of sample size on the accuracy of estimates of unknown parameters, we take $N = 300$ and 800, which represent the small and the large levels of sample sizes.

For each data set we consider two Bayesian estimates for the unknown parameters $\boldsymbol{\theta}$: one is obtained under the parametric setting (denoted by 'BAY I') and the other is under the semiparametric setting (denoted by 'BAY II'). The parametric method is defined as Equations (S.24) associated with (S.23), while the semiparametric model is specified via Equations (S.6) with truncated level $G = 100$. Note that under the parametric setting, Equations (S.4) and the prior in the main text together specify the priors of model parameters. The following inputs of the hyperparameters in (S.4) and the prior distributions are considered both for obtaining estimates BAY I and II: $\alpha_{10} = 0$, $\alpha_{20} = 0.00$, $A_{10} = 100$, $A_{20} = 100$, $\mu_{0k} = 0.00$, $A_{\mu 0k} = 100.0(k = 1, \cdots, 6)$, $\alpha_{\epsilon 0} = \alpha_{\delta k 0} = \alpha_{\zeta \ell 0} = 9.0$, $\beta_{\epsilon 0} = \beta_{\delta k 0} = \beta_{\zeta \ell 0} = 8.0$, $\widetilde{\boldsymbol{\gamma}}_{10} = \widetilde{\boldsymbol{\gamma}}_{20} = \mathbf{0}_6$, $\widetilde{\mathbf{\Sigma}}_{1j0} = \widetilde{\mathbf{\Sigma}}_{2j0} = 100 \times \mathbf{I}_6$, $\mathbf{\Lambda}_{0k}$ and $\mathbf{\Pi}_{j0}$ are set to be the true values in $\mathbf{\Lambda}$ and $\mathbf{\Pi}$; $\mathbf{H}_{0\delta j} = \mathbf{I}_3$, $H_\zeta = 1.0$, $\rho_0 = 10.0$ and $\mathbf{R}_0^{-1} = 7\mathbf{I}_2$. These values are the standard inputs ensuring the priors to be inflated enough. In addition, we take $\lambda_0 = \nu_0 = 2.0$ to encourage smaller and larger values for c. We also try other inputs for the hyperparameters and find thatthe resulting estimates are more robust against these choices.

The proposed MCMC algorithm given in Section S2 is implemented to obtain estimates of unknown parameters. For the MH algorithm, the values of tuning parameters are taken to keep the acceptance rate about 0.43 in all scenarios. A few test runs are conducted as pilots to monitor the convergence of the Gibbs sampler. We plot the values of EPSR (Gelman and Rubin, 1987) of unknown parameters against the number of iterations under three dispersed starting values. Figure 1 presents the plots of values of EPSR under the semiparametric setting for the 'MIX' data with

$N = 300$ (the others are similar and omitted for saving space). It can be found that

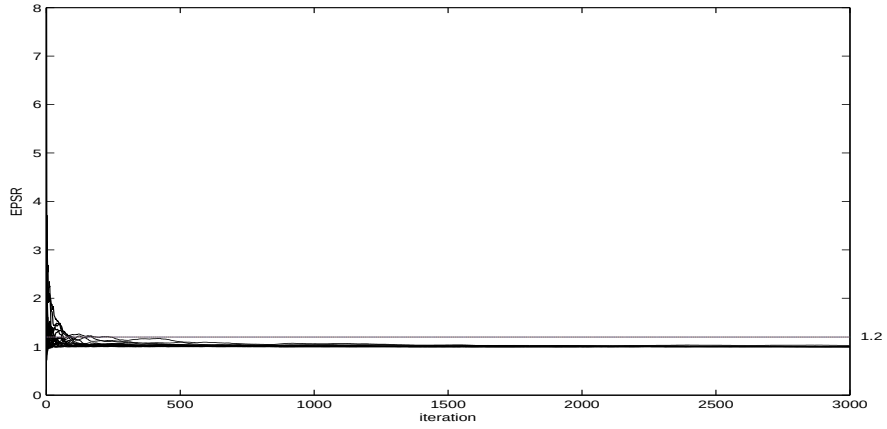

Figure 1: Plots of values of EPSR of estimates of unknown parameters against the number of iterations under three different staring values in the simulation study: semiparametric model with $G = 100$ and $N = 300$

.

the convergence of estimates is fast and the values of EPSR are less than 1.2 in about 800 iterations. To be conservative, we remove the first 2000 observations as burn-in and collect 3000 observations further for computing the bias (BIAS), the root mean squares (RMS) and the standard deviation (SD) across 100 replications. The BIAS and RMS of the j-th component $\widehat{\theta}_j$ in estimates are defined as follows:

$$\text{BIAS}(\widehat{\theta}_j) = (\bar{\theta}_j - \theta_j^0), \bar{\theta}_j = \frac{1}{100}\sum_{\kappa=1}^{100}\widehat{\theta}_j^{(\kappa)}, \quad \text{RMS}(\widehat{\theta}_j) = \sqrt{\frac{1}{100}\sum_{\kappa=1}^{100}(\widehat{\theta}_j^{(\kappa)} - \theta_j^0)^2}, \text{ (S.28)}$$

where θ_j^0 is the j-th element of population parameters $\boldsymbol{\theta}^0$. The summaries of estimates under two fittings are reported in Tables 1 to 3, where the sums of the SD and RMS across the estimates are presented in the last rows.

Examinations of Tables 1 to 3 present the following findings: (i) For the 'MIX' and 'HT' data, the estimate BAY I produces incorrect results. For 'MIX' data, the total RMS and SD are 10.727 and 4.106 for $N = 300$, and 10.505 and 2.591 for $N = 800$ respectively, while for 'HT' data, the sum of RMS and the sum of SD are respectively 13.504 and 5.888 for $N = 300$ and 13.711 and 4.12 for $N = 800$. This

Table 1: Summary statistics of estimates of unknown parameters based on the 'PAR' data: simulation study.

	$N = 300$						$N = 800$					
	BAY I			BAY II			BAY I			BAY II		
Para.	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD
γ_{11}	-0.102	0.355	0.385	0.039	0.467	0.518	-0.022	0.236	0.238	-0.173	0.381	0.354
γ_{12}	-0.141	0.299	0.553	0.106	0.199	0.207	-0.087	0.442	0.379	0.036	0.134	0.126
γ_{13}	0.167	0.299	0.228	-0.163	0.429	0.401	0.020	0.160	0.129	-0.044	0.306	0.249
β_{11}	0.076	0.259	0.255	0.045	0.337	0.250	0.027	0.131	0.146	0.039	0.154	0.149
β_{12}	0.197	0.383	0.347	0.217	0.428	0.337	0.056	0.208	0.194	0.097	0.207	0.195
β_{13}	0.134	0.285	0.341	0.121	0.534	0.345	0.029	0.185	0.191	0.011	0.158	0.192
γ_{21}	-0.069	0.218	0.207	-0.02	0.256	0.324	-0.016	0.115	0.124	-0.054	0.199	0.202
γ_{22}	-0.165	0.339	0.346	-0.014	0.080	0.103	-0.059	0.232	0.208	0.019	0.052	0.062
γ_{23}	0.018	0.091	0.112	-0.114	0.194	0.218	0.000	0.043	0.064	0.037	0.115	0.127
β_{21}	0.052	0.116	0.116	0.014	0.109	0.111	0.002	0.055	0.067	-0.002	0.059	0.066
β_{22}	-0.066	0.208	0.203	-0.029	0.189	0.201	-0.024	0.112	0.117	0.008	0.090	0.115
β_{23}	-0.057	0.172	0.201	-0.051	0.155	0.203	-0.032	0.100	0.118	0.034	0.102	0.119
Λ_{22}	0.000	0.028	0.032	0.008	0.026	0.032	0.000	0.019	0.020	0.004	0.021	0.019
Λ_{43}	-0.005	0.032	0.057	-0.001	0.044	0.056	0.009	0.036	0.035	0.005	0.027	0.035
Λ_{53}	0.008	0.048	0.056	0.007	0.053	0.058	-0.012	0.043	0.035	-0.002	0.030	0.035
Π_{11}	0.040	0.104	0.108	0.042	0.088	0.110	0.046	0.075	0.064	0.025	0.055	0.063
Π_{12}	0.035	0.114	0.106	0.054	0.117	0.108	-0.006	0.081	0.066	0.033	0.066	0.065
Π_{13}	0.033	0.115	0.101	0.064	0.100	0.104	0.011	0.065	0.058	0.020	0.064	0.060
ψ_ζ	-0.136	0.191	0.133	-0.108	0.146	0.135	-0.040	0.098	0.091	-0.073	0.105	0.089
Φ_{22}	-0.053	0.084	0.060	-0.074	0.093	0.060	-0.055	0.066	0.037	-0.059	0.066	0.037
Total	-	3.741	3.946	-	4.043	3.882	-	2.502	2.381	-	2.392	2.360

Table 2: Summary statistics of estimates of unknown parameters based on the 'MIX' data: simulation study.

	$N = 300$						$N = 800$					
	BAY I			BAY II			BAY I			BAY II		
Para.	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD
γ_{11}	0.093	0.388	0.538	0.119	0.364	0.566	-0.122	0.363	0.372	-0.173	0.381	0.354
γ_{12}	-0.071	0.195	0.198	0.048	0.219	0.222	-0.119	0.192	0.129	0.036	0.134	0.126
γ_{13}	-0.099	0.287	0.421	0.049	0.413	0.458	-0.015	0.192	0.274	-0.044	0.306	0.249
β_{11}	0.085	0.247	0.254	0.009	0.202	0.265	0.131	0.244	0.163	0.039	0.154	0.149
β_{12}	-0.911	0.954	0.327	0.148	0.349	0.342	-0.998	1.016	0.197	0.097	0.207	0.195
β_{13}	-0.963	0.996	0.322	-0.112	0.244	0.324	-1.022	1.047	0.195	0.011	0.158	0.192
γ_{21}	0.106	0.306	0.349	0.057	0.293	0.291	0.011	0.185	0.232	-0.054	0.199	0.202
γ_{22}	0.003	0.107	0.110	-0.028	0.076	0.093	-0.017	0.049	0.071	0.019	0.052	0.062
γ_{23}	0.020	0.183	0.247	0.053	0.237	0.203	-0.116	0.232	0.155	0.037	0.115	0.127
β_{21}	0.639	0.651	0.134	0.014	0.073	0.107	0.582	0.588	0.077	-0.002	0.059	0.066
β_{22}	-1.280	1.291	0.214	-0.005	0.134	0.175	-1.205	1.214	0.123	0.008	0.090	0.115
β_{23}	-1.256	1.273	0.206	-0.057	0.151	0.170	-1.272	1.279	0.121	0.034	0.102	0.119
Λ_{22}	0.127	0.128	0.030	0.008	0.037	0.041	0.121	0.121	0.018	0.004	0.021	0.019
Λ_{43}	0.910	0.912	0.066	-0.015	0.066	0.072	0.930	0.930	0.040	0.005	0.027	0.035
Λ_{53}	0.898	0.900	0.068	-0.026	0.079	0.071	0.928	0.928	0.040	-0.002	0.030	0.035
Π_{11}	0.194	0.251	0.145	0.023	0.125	0.122	0.176	0.197	0.084	0.025	0.055	0.063
Π_{12}	0.163	0.237	0.140	0.000	0.120	0.121	0.176	0.200	0.084	0.033	0.066	0.065
Π_{13}	-0.655	0.660	0.064	0.018	0.105	0.118	-0.665	0.667	0.043	0.020	0.064	0.060
ψ_ζ	0.621	0.668	0.220	-0.120	0.187	0.156	0.746	0.759	0.141	-0.073	0.105	0.089
Φ_{22}	0.078	0.092	0.052	-0.069	0.083	0.063	0.099	0.100	0.032	-0.059	0.066	0.037
Total	-	10.727	4.106	-	3.556	3.978	-	10.505	2.591	-	2.392	2.360

Table 3: Summary statistics of estimates of unknown parameters based on the 'HT' data: simulation study.

	$N = 300$						$N = 800$					
	BAY I			BAY II			BAY I			BAY II		
Para.	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD	BIAS	RMS	SD
γ_{11}	-0.152	0.384	0.39	-0.261	0.204	0.207	-0.379	0.566	0.535	-0.209	0.150	0.156
γ_{12}	0.049	0.176	0.154	0.062	0.229	0.212	0.058	0.25	0.222	0.064	0.156	0.136
γ_{13}	0.033	0.273	0.281	-0.038	0.295	0.400	0.059	0.365	0.411	0.004	0.225	0.154
β_{11}	-0.129	0.793	0.224	0.106	0.207	0.176	-0.127	0.856	0.292	0.057	0.158	0.163
β_{12}	0.462	0.819	0.320	0.127	0.244	0.177	0.276	0.913	0.473	0.133	0.155	0.128
β_{13}	0.550	0.993	0.339	0.152	0.227	0.172	0.199	0.705	0.434	0.182	0.204	0.135
γ_{21}	-0.202	0.374	0.372	-0.248	0.288	0.201	-0.281	0.473	0.540	-0.132	0.206	0.166
γ_{22}	-0.007	0.123	0.132	-0.015	0.150	0.138	-0.059	0.225	0.203	0.012	0.101	0.086
γ_{23}	-0.059	0.241	0.257	-0.114	0.285	0.290	0.037	0.350	0.389	-0.046	0.187	0.173
β_{21}	-0.077	1.076	0.201	-0.029	0.195	0.192	-0.040	0.679	0.260	-0.028	0.122	0.107
β_{22}	0.281	1.261	0.347	0.060	0.242	0.253	0.238	1.293	0.521	0.043	0.214	0.193
β_{23}	0.576	2.279	0.383	0.076	0.254	0.238	0.140	1.253	0.461	0.087	0.232	0.200
Λ_{22}	0.190	0.775	0.083	0.009	0.046	0.043	-0.297	1.113	0.081	0.004	0.025	0.025
Λ_{43}	0.078	1.368	0.103	0.018	0.071	0.075	-0.141	1.360	0.163	0.011	0.040	0.044
Λ_{53}	0.266	0.859	0.101	0.018	0.090	0.076	0.062	0.908	0.162	0.018	0.046	0.044
Π_{11}	-0.003	0.280	0.106	0.082	0.150	0.126	0.082	0.496	0.181	0.019	0.071	0.080
Π_{12}	0.140	0.339	0.100	0.040	0.153	0.133	-0.082	0.399	0.176	0.038	0.092	0.079
Π_{13}	0.397	0.536	0.103	0.059	0.170	0.134	0.296	0.666	0.154	0.032	0.079	0.078
ψ_ζ	-0.564	0.584	0.080	-0.114	0.196	0.166	-0.311	0.408	0.164	-0.074	0.136	0.115
Φ_{22}	-0.149	0.179	0.044	-0.07	0.089	0.066	-0.167	0.226	0.068	-0.059	0.071	0.040
Total	-	13.711	4.120	-	3.785	3.475	-	13.504	5.888	-	2.670	2.302

shows that that the usual parametric fitting is more sensitive to the distributional deviations; However, the estimate BAY II is more reasonable. The total RMS and SD are just 3.556 and 3.978 even for small sample size $N = 300$. This indicates that the semiparametric modeling is more robust against distributional deviations; (ii) For the 'PAR' data, two estimates both give the reasonable results. The estimate BAY I is more accurate than BAY II but the differences between them are slight. This is not surprising since in this case, the parametric fitting is consistent with the true population model; (iii) As expected, increasing the sample size improves the accuracy of two estimates no matter whether the postulated model is specified correctly or not.

In the following, we focus on assessing the performance of L_ν in selecting competing models under consideration. Two set-ups for the true population model are considered: one is the parametric model given by (S.24) (denoted by 'PAR') and the other is the mixture model given by (S.25)(denoted by 'MIX'). The sample size is taken as $N = 300$. In view of the key role of latent factors in the current analysis, we mainly concentrate upon exploring their effects on the different parts of the two-part model. In addition, we also focus on determining whether or not the interaction is proper for structural equation. Hence, the following competing models are considered:

M_1 : Full model with $(\beta_{11}, \beta_{12}, \beta_{13}) = \mathbf{0}$ in Part one;

M_2 : Full model with $(\beta_{21}, \beta_{22}, \beta_{23}) = \mathbf{0}$ in Part two;

M_3 : Full model with $(\beta_{11}, \beta_{12}, \beta_{13}) = (\beta_{21}, \beta_{12}, \beta_{13}) = \mathbf{0}$ in Parts one and two;

M_4 : Full model with $\Gamma_3 = 0$ in $(S.23)$.

Here, the overall model without above constraints is referred to as the 'Full model' (denoted by M_0). Obviously, M_1 to M_4 represent different levels of deviations away from the true model. Among them, M_1 and M_2 are that the effects of latent factors are not imposed on the discrete and continuous parts respectively, while M_3 indicates that the two-part model and the measurement model are separated; M_4 refers to the linear structure for causal effects of the exogenous on the endogenous factors. Note that these models may not be nested since some elements in $\boldsymbol{\theta}$ are fixed previously for model determination or for identifying the scales of factors.

For each data set, we calculate $L_{0.5}$ across $M_j(j = 0, 1, \cdots, 4)$ under the parametric (denoted by 'Para') and the semiparametric (denoted by 'Semi') settings respectively. Moreover, for comparison, we also calculate $L_{0.1}$ and L_1 in all scenarios. Note that the measure L_1 was originally proposed by Ibrahim and Laud (1994) in Bayesian model comparison and not a special case of L_ν. It is also noted that for the 'PAR' data, M_0 under the parametric setting corresponds to the true model. However, no competing models are consistent with the true population model for the 'MIX' data. For computation, since the unknown parameters of competing models are not identical, the priors to these parameters should be assigned on the model-to-model basis. To tackle such problem, we first assign the priors to the parameters involved in the overall model, i.e., the full model without constraints. Then, the priors of parameters under other fittings are obtained via transformations. For example, the prior of (Γ_1, Γ_2) can be obtained from the distribution of $(\Gamma_1, \Gamma_2, \Gamma_3)$ via transformation $(\Gamma_1, \Gamma_2)^T = A(\Gamma_1, \Gamma_2, \Gamma_3)^T$ where A is the 2×3 permutation matrix. For each fitting, we implement the MCMC sampling algorithm and collect 3000 observations after the first 2000 iterations to compute $L_{0.1}$, $L_{0.5}$ and L_1. The resulting summaries of estimates across 100 replications are given in Tables 4 to 6.

It follows from Tables (4) and (6) that: (i) For the 'PAR' data, all measures give the correct selection regardless of the parametric and the semi-parametric fittings. They both attain the minimums at M_0. Under the parametric setting, the minimum values of $L_{0.1}$, $L_{0.5}$ and L_1 equal to 946.400, 1031.256, 1137.362 with SD=10.396, 51.987 and 76.600 respectively, while under the semiparametric setting, the minimum values of $L_{0.1}$, $L_{0.5}$ and L_1 are 958.169,1132.072,1349.451 with SD=19.046, 95.231, 120.364. It also follows that at M_0 all L-measures under parametric fitting are smaller than those under semiparametric fitting. This agrees with the situations where the data set is generated from the parametric model; (ii) For the 'MIX' data, the situation becomes more complex since no competing models under consideration are consistent with the true model. We can not tell about which one outperforms the other. But based on our simulation study, the minimums of all measures favor M_0 under the semiparametric fitting. The potential reason can be attributed to the fact

Table 4: Summary of estimates of $L_{0.1}$-measure under various competing models in the simulation study: $N = 300$ and $G = 100$.

		$L_{0.1}$-measure			
		Para		Semi	
Data	Model	Est.	SD	Est.	SD
PAR					
	M_0	946.400	10.396	958.169	19.046
	M_1	958.352	9.270	976.854	16.185
	M_2	1350.910	49.321	1308.453	31.160
	M_3	1419.084	38.128	1276.791	62.627
	M_4	961.574	26.662	986.390	26.663
MIX					
	M_0	2787.770	91.107	1970.035	29.571
	M_1	2877.986	155.410	2076.927	42.546
	M_2	3986.680	73.036	4129.568	54.266
	M_3	4136.948	71.012	4297.136	96.848
	M_4	2833.276	51.333	2026.101	73.351

Table 5: Summary of estimates of $L_{0.5}$-measure under various competing models in the simulation study: $N = 300$ and $G = 100$.

		$L_{0.5}$-measure			
		Para		Semi	
Data	Model	Est.	SD	Est.	SD
PAR					
	M_0	1031.256	51.978	1132.072	95.231
	M_1	1078.415	46.351	1182.391	80.929
	M_2	2150.235	246.605	1623.083	155.802
	M_3	1640.884	190.641	1868.610	313.135
	M_4	1089.929	133.309	1224.291	133.316
MIX					
	M_0	3134.104	455.537	3000.443	147.855
	M_1	3297.677	777.049	3113.190	212.728
	M_2	4426.363	365.178	4569.929	271.328
	M_3	4706.951	355.058	4796.253	484.240
	M_4	3137.173	256.667	3074.718	366.757

Table 6: Summary of estimates of L_1-measure under various competing models in the simulation study: $N = 300$ and $G = 100$.

		L_1-measure			
		Para		Semi	
Data	Model	Est.	SD	Est.	SD
PAR					
	M_0	1137.326	76.600	1349.451	120.364
	M_1	1228.493	25.520	1439.312	106.675
	M_2	3149.391	135.107	3316.371	175.211
	M_3	1918.134	128.301	2008.382	144.641
	M_4	1250.373	125.601	1521.667	138.340
MIX					
	M_0	4288.453	155.204	3567.022	161.301
	M_1	4408.519	252.121	3822.291	252.941
	M_2	5120.380	201.804	5175.967	293.981
	M_3	5420.149	425.501	5419.455	482.407
	M_4	4385.489	333.101	3517.044	374.401

that the semiparametric model is much closer to the true model when compared to the parametric model; (iii) We also find that for 'MIX' data, M_2 and M_3 provide poor fits when compared to other competing models. The underlying reason perhaps is that within our framework, the latent factors have significant influences on the continuous parts, and the absence of them will seriously distort the conclusion;

For computation, all programs are coded in C language and implemented on Inter(R) Core(TM), i5-6500 processor with CPU 3.20GHz on Microsoft Windows 7 operating system. In our simulation study, the convergence is fast and every 5000 iterations in blocked Gibbs sampler under the semi-parameter setting only needs 43.61 seconds for $N = 800$ and $G = 100$. It takes about six hours to complete 100 replications. Request for codes can be send to the corresponding author.

References

[1] Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

[2] Cowles, M. K. (1996). Accelerating Monte Carlo Markov Chains convergence for cumulative-link generalized linear modes. *Statistics and Computing*, **6**, 101-111.

[3] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**(2), 209-230.

[4] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**,457–511.

[5] Gilks, W. R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **41**, 337-348.

[6] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97-109.

[7] Holmes, C. C. and Held, L.(2006). Bayesian auxiliary variable models for binary and multinominal regression. *Bayesian Analysis*, **1**, 145-168.

[8] Ibrahim, J. G. and Laud, P. W. (1994). A predictive approach to the analysis of designed experiments. *Journal of the American Statistical Association*, **89**(430), 309-319.

[9] Ishwaran, H. and James, L. F. (2001). Gibbs Sampling Methods for stick-breaking priors. *Journal of the American Statistical Association*, **96**, 161-173.

[10] Ishwaran, H., & Zarepour, M. (2000). Markov chain Monte Carlo in approximation Dirichlet and beta-parameter process hierarchical models. *Biometrika*, **87**, 371-390.

[11] Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International: Hove and London.

[12] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, **21**, 1087-1091.

[13] Polson, N .G., Scott, J. G. and Windle, J. (2013). Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables. *Journal of the American Statistical Association*, **108**, 1339-1349.

[14] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. Statistica Sinica, **4**, 639C650.

[15] Walker, S. (2007). Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics-Simulation and Computation*,**36**, 45-54.

[16] Xia, Y. and Gou, J. W. (2016). Bayesian semiparametric analysis for latent variable models with mixed continuous and ordinal outcomes. *Journal of the Korean Statistical Society*, **45**, 451-465.