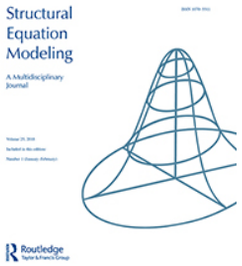
Routledge
Taylor & Francis Group



Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

# Bayesian Analysis of Mixture Structural Equation Models With an Unknown Number of Components

Hefei Liu & Xin Yuan Song

**To cite this article:** Hefei Liu & Xin Yuan Song (2018) Bayesian Analysis of Mixture Structural Equation Models With an Unknown Number of Components, Structural Equation Modeling: A Multidisciplinary Journal, 25:1, 41-55, DOI: 10.1080/10705511.2017.1372688

**To link to this article:** https://doi.org/10.1080/10705511.2017.1372688

*Structural Equation Modeling: A Multidisciplinary Journal*, 25: 41–55, 2018

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2017.1372688



# Bayesian Analysis of Mixture Structural Equation Models With an Unknown Number of Components

Hefei Liu[1] and Xin Yuan Song[2]
[1]*Qujing Normal University*
[2]*The Chinese University of Hong Kong*

Multivariate heterogenous data with latent variables are common in many fields such as biological, medical, behavioral, and social-psychological sciences. Mixture structural equation models are multivariate techniques used to examine heterogeneous interrelationships among latent variables. In the analysis of mixture models, determination of the number of mixture components is always an important and challenging issue. This article aims to develop a full Bayesian approach with the use of reversible jump Markov chain Monte Carlo method to analyze mixture structural equation models with an unknown number of components. The proposed procedure can simultaneously and efficiently select the number of mixture components and conduct parameter estimation. Simulation studies show the satisfactory empirical performance of the method. The proposed method is applied to study risk factors of osteoporotic fractures in older people.

**Keywords**: Bayesian method, latent variables, mixture models, RJMCMC algorithm

As an analytic tool deeply rooted in medical, behavioral, psychological, and social sciences, structural equation models (SEMs) are one of the most widely used methods for assessing plausible causal assumptions by modeling relationships between latent variables (Bentler & Wu, 2004; Bollen, 1989; Jöreskog & Sörbom, 1996; Lee, 2007; Song & Lee, 2012). In the analysis of SEMs, researchers often treat the data as if they were sampled from a single population. Given that this homogeneity assumption might be unrealistic in substantive research and a violation of it will lead to erroneous inference results (e.g., Jedidi, Jagpal, & Desarbo, 1997), appropriately addressing heterogeneity has become a major issue and has received much attention in the literature. If heterogeneity is present and the component membership is known beforehand, then standard multiple group methods can be applied without difficulty. However, component membership is usually unknown in practice. Mixture SEMs with an unknown number of components perfectly accommodate the unobservable heterogeneity and reveal the component-specific interrelationships among latent variables.

Mixture modeling is an important tool for analyzing heterogeneous data in many contexts (e.g., Richardson & Green, 1997; Roeder & Wasserman, 1997; Titterington, Smith, & Markov, 1985). Various methods have been proposed to estimate mixture models with a fixed number of components (e.g., Diebolt & Robert, 1994; Lindsay & Basak, 1993). For mixture models with an unknown number of components, the main issues involved typically include the determination of the number of components, the estimation of the component membership for each individual, and the component-specific model parameters. In the context of mixture SEMs, Jedidi et al. (1997) analyzed a finite mixture of SEMs using a maximum likelihood (ML) approach; in this method, the expectation-maximization (EM) type algorithm was implemented to obtain the ML estimates of parameters, and the Bayesian information criterion (BIC) was applied to determine the number of mixture components. Yung (1997) investigated a mixture of confirmatory factor analysis models and proposed an approximated scoring algorithm copied with an EM algorithm to conduct estimation. Dolan and Van Der Mass (1998) used a quasi-Newton algorithm and inferred the estimation by changing the degrees of separation and the sample sizes. Arminger,

Correspondence should be addressed to Xin Yuan Song, Department of Statistics, Chinese University of Hong Kong, Shatin, Hong Kong. E-mail: xysong@sta.cuhk.edu.hk

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HSEM.

Stein, and Witterberg (1999) proposed three estimation strategies for mixtures of conditional mean and covariance structure models. Zhu and Lee (2001) proposed a Bayesian approach coupled with Markov chain Monte Carlo (MCMC) methods to analyze a finite mixture of LISREL models. Lee and Song (2003) employed Bayes factor to select the number of components in mixture SEMs through the path sampling procedure. Cai and Song (2010) and Cai, Song, and Hser (2010) proposed a modified deviance information criterion to determine the number of components for mixture SEMs in the presence of missing data. Li, Yutaka, Pan, and Song (2012) considered the use of $L$-measure to compare mixture SEMs with different numbers of components. Nevertheless, the aforementioned methods are criterion- or classical test statistic-based and used to determine the number of mixture models in a pairwise basis; hence, these methods become increasingly tedious when the model space becomes large. Moreover, published reports (Celeux, Forbes, Robert, & Titterington, 2006; Plummer, 2008; Spiegelhalter, Thomas, Best, & Lunn, 2003) argued that criterion-based statistics tend to select complex models with overestimated numbers of mixture components.

Green (1995) proposed the reversible jump MCMC (RJMCMC) method, which regards the model dimension as unknown and estimates it together with other parameters. Later on, the RJMCMC method has been applied to analyze various mixture or dynamic mixture models (e.g., Papastamoulis & Iliopoulos, 2009; Richardson & Green, 1997; Robert, Rydén, & Titterington, 2000). However, few works have focused on a simultaneous component selection and parameter estimation in the context of mixture SEMs. This article aims to develop a full Bayesian approach for analysis of mixture SEMs with an unknown number of components. We develop a hybrid algorithm, which combines the Gibbs sampler (Geman & Geman, 1984), RJMCMC algorithm, and the Metropolis–Hastings (MH; Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) algorithm, to conduct posterior inference. Given the existence of latent variables and complexity of the proposed model, the split (or merge) step of some parameters (e.g., the covariance matrix of the latent variables) is challenging. We propose novel split (or merge) strategies based on a Cholesky decomposition principle to address the difficulty. In contrast to existing criterion- or test-statistic-based approaches that must predetermine the number of mixture components before estimation through a tedious pairwise model comparison, the proposed procedure allows the number of mixture components to vary across MCMC iterations and estimates it according to the posterior probabilities of all candidate models. With the use of the RJMCMC procedure, all the unknowns including the number of mixture components, individuals' component memberships, and component-specific model parameters can be estimated together. As shown by the simulation of this study, the proposed procedure facilitates a fast exploration in a large model space and is therefore much more computationally efficient than available criterion-based methods.

A number of researchers (e.g., Fox, Sudderth, Jordan, & Willsky, 2008; Song, Xia, Pan, & Lee, 2011; Teh, Jordan, Beal, & Blei, 2006; Yang & Dunson, 2010) applied centered or truncated Dirichlet processes for mixture modeling (e.g., Ishwaran & James, 2001; Ishwaran & Zarepour, 2000) to automatically select finite components from infinite mixtures; however, these approaches are mainly developed to approximate the unknown distributions of latent variables or random errors but cannot be immediately used to determine the number of components and examine the heterogeneous interrelationships among latent variables in mixture SEMs. The use of recently developed regularized methods is another promising direction of conducting simultaneous component selection and parameter estimation in mixture models. For instance, Chen and Khalili (2008) proposed a penalized likelihood method with the Smoothly Clipped Absolute Deviation (SCAD) penalty on the mixing proportions and the component parameters to conduct order selection in finite mixture models. Huang, Peng, and Zhang (2017) proposed a modified EM algorithm, which penalizes the logarithm of mixing proportions in multivariate Gaussian mixture models and simultaneously deletes components with the mixing proportions being shrunk to zero. However, these methods only considered simple Gaussian mixtures without the complex covariance structure described by SEMs and were all developed in the ML framework. Despite recent advances in Bayesian regularized procedures (e.g., Feng, Wu, & Song, 2017a, 2017b; Guo, Zhu, Chow, & Ibrahim, 2012), the existing developments emphasized variable selection in regression. For mixture models, how to introduce penalty to mixing proportion parameters by assigning appropriate prior distributions and how to conduct the associated posterior inference are highly challenging. Thus, the feasibility of conducting a Bayesian regularized procedure for mixture SEMs remains unknown and requires further investigation.

The article is organized as follows. The next section defines mixture SEMs. The section after that introduces the Bayesian inference procedure and describes implementation of the RJMCMC algorithm in the context of mixture SEMs. We then present simulation studies to evaluate the empirical performance of the proposed method. Finally, we report an application to a study of the risk factors of osteoporotic fractures in older people before concluding the article.

## MIXTURE OF STRUCTURAL EQUATION MODELS

Let $\boldsymbol{y}_i$ be a $p \times 1$ random vector. A mixture of SEMs is defined as

$$f(\boldsymbol{y}_i) = \sum_{k=1}^{K} \pi_k f_k(\boldsymbol{y}_i | \boldsymbol{\tau}_k), \quad i = 1, \cdots, n, \tag{1}$$

where $K$ is the number of components, $\pi_k$ is the unknown mixing proportion such that $\pi_k > 0$ and $\pi_1 + \cdots + \pi_K = 1$, $f_k(\boldsymbol{y}_i|\ \boldsymbol{\tau}_k)$ is a multivariate normal density function, and $\boldsymbol{\tau}_k$ is a vector of the unknown parameters of component $k$.

Within the $k$th component of the mixture, we specify an SEM as follows:

$$\boldsymbol{y}_i = \boldsymbol{\mu}_k + \boldsymbol{\Lambda}_k \boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, \tag{2}$$

$$\boldsymbol{\eta}_i = \prod_k \boldsymbol{\eta}_i + \boldsymbol{\Gamma}_k \boldsymbol{\xi}_i + \boldsymbol{\delta}_i = \boldsymbol{\Lambda}_{\omega k} \boldsymbol{\omega}_i + \boldsymbol{\delta}_i, \tag{3}$$

where $\boldsymbol{\mu}_k$ is a $p \times 1$ intercept vector, $\boldsymbol{\Lambda}_k$ is a $p \times q$ factor loading matrix, $\prod_k$ and $\boldsymbol{\Gamma}_k$ are $q_1 \times q_1$ and $q_1 \times q_2$ matrices of regression coefficients, respectively, $\boldsymbol{\Lambda}_{\omega k} = (\prod_k,\ \boldsymbol{\Gamma}_k)$; $\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$ is a $q \times 1$ vector of latent variables, $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$ are $q_1 \times 1$ and $q_2 \times 1$ vectors of outcome and explanatory latent variables, respectively, and $\boldsymbol{\xi}_i \sim N(\mathbf{0},\ \boldsymbol{\Phi}_k)$; $\boldsymbol{\varepsilon}_i$ is a $p \times 1$ vector of random errors independent of $\omega_i$ and distribution as $N(\mathbf{0},\ \boldsymbol{\Psi}_k)$ with diagonal $\boldsymbol{\Psi}_k$, and $\boldsymbol{\delta}_i$ is a $q_1 \times 1$ vector of random errors independent of $\boldsymbol{\xi}_i$ and distributed as $N(\mathbf{0},\ \boldsymbol{\Psi}_{\delta k})$ with diagonal $\boldsymbol{\Psi}_{\delta k}$.

The number of latent variables $q$ and the structure of factor loading matrix $\boldsymbol{\Lambda}_k$ in Equation 2 can be naturally decided based on the study objectives, the meaning of the observed variables suggested by subject experts, or the existing literature. For instance, in the osteoporosis study presented later, medical experts and previous studies (e.g., Lau et al., 2006) suggested that bone mineral density (BMD), obesity, estrogen, androgen, precursors, and metabolites should be characterized by observed variables {spine BMD, hip BMD}, {BMI, Waist hip ratio}, {E1, E1-S, E2}, {TESTO, 5-DIOL, DHT}, {4-DIONE, DHEA, DHEA-S}, and {ADT, ADT-G, 3G, 3G-17G}, respectively, and no observed variable characterizes more than one latent variable. Thus, $p$ = 17, $q$ = 6, and $\boldsymbol{\Lambda}_k$ has a nonoverlapping structure of Equation 8. If such information is unavailable, then an exploratory factor analysis (EFA) can be used to determine $p$, $q$, and the structure of $\boldsymbol{\Lambda}_k$ based on the data. Well-known software in SEMs, such as LISREL (Jöreskog & Sörbom, 1996), EQS (Bentler & Wu, 2004), M*plus* (Muthén & Muthén, 1998–2007), and AMOS can be used to conduct the EFA.

The mixture SEMs defined by Equations 1 through 3 is not identifiable. The first indeterminacy is label switching owing to the invariance of the likelihood function to a random permutation of component labels, leading to a multimodal posterior distribution under symmetric prior. Many methods have been proposed to address this label switching problem (Boys & Henderson, 2001; Frühwirth-Schnatter, 2001; Jasra, Holmes, & Stephens, 2005; Matthew, 2000; Papastamoulis & Iliopoulos, 2009; Richardson & Green, 1997; Spezia, 2009). In this study, we simply use the constraint $\mu_{1,1} < \mu_{2,1} < \cdots < \mu_{K,1}$ to identify the model, where $\mu_{k,1}$ is the first element of $\boldsymbol{\mu}_k$, $k = 1, \cdots, K$. The second indeterminacy is caused by the multiplication of two unknown parameters, $\boldsymbol{\Lambda}_k\ \boldsymbol{\omega}_i$, in Equation 2. We follow a common practice in the latent variable modeling literature (Song & Lee, 2012) to fix appropriate elements in $\boldsymbol{\Lambda}_k$ to solve the problem.

## BAYESIAN INFERENCE

This section consists of three parts. The first part specifies the prior distributions of the unknown parameters. The second part discusses the posterior sampling and the associated full conditional distributions. The last part describes the RJMCMC algorithm and its implementation in the context of mixture SEMs.

### Prior Specification

The first task of a full Bayesian analysis is to specify appropriate prior distributions for the unknown model parameters. On the basis of the existing literature (Lee & Song, 2003; Richardson & Green, 1997), we consider the following prior distributions.

For $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_K)^T$, we assign a symmetric Dirichlet distribution as follows:

$$\boldsymbol{\pi} \overset{D}{=} D(\alpha, \cdots, \alpha), \tag{4}$$

where $\alpha$ is a hyperparameter with a prespecified value.

Let $\boldsymbol{\Lambda}_{\omega k} = (\prod_k,\ \boldsymbol{\Gamma}_k)$, $\boldsymbol{\Lambda}_{km}^T$ and $\boldsymbol{\Lambda}_{\omega kl}^T$ be the $m$th and $l$th rows of $\boldsymbol{\Lambda}_k$ and $\boldsymbol{\Lambda}_{\omega k}$, respectively, and $\psi_{km}$ and $\psi_{\delta kl}$ be the $m$th and $l$th diagonal element of $\boldsymbol{\Psi}_k$ and $\boldsymbol{\Psi}_{\delta k}$, respectively. For the structural parameters involved in Equations 2 and 3, we assign conjugate prior distributions as follows (Arminger & Muthen, 1998; Shi & Lee, 1998; Song & Lee, 2012): for $m = 1, \cdots, p$ and $l = 1, \cdots, q_1$,

$$\begin{aligned}
[\boldsymbol{\Lambda}_{km}|\psi_{km}] &\overset{D}{=} N[\boldsymbol{\Lambda}_{0km}, \psi_{km}\mathbf{H}_{0ykm}], \psi_{km}^{-1} \overset{D}{=} Gamma[\alpha_{0\varepsilon k}, \beta_{0\varepsilon k}], \\
[\boldsymbol{\Lambda}_{km}|\psi_{km}] &\overset{D}{=} N[\boldsymbol{\Lambda}_{0wkl}, \psi_{\delta kl}\mathbf{H}_{0\omega kl}], \psi_{\delta kl}^{-1} \overset{D}{=} Gamma[\alpha_{0\delta k}, \beta_{0\delta k}], \\
\boldsymbol{\mu}_k &\overset{D}{=} N[\boldsymbol{\mu}_{0k}, \textstyle\sum_{0k}], \boldsymbol{\Phi}_k^{-1} \overset{D}{=} W_{q_2}[\mathbf{R}_{0k}, \rho_{0k}],
\end{aligned} \tag{5}$$

where $\boldsymbol{\Lambda}_{0km}, \alpha_{0\varepsilon k}, \beta_{0\varepsilon k}$, $\boldsymbol{\Lambda}_{0\omega kl}, \alpha_{0\delta k}, \beta_{0\delta k}$, $\boldsymbol{\mu}_{0k}, \rho_{0k}$, and positive definite matrices $\mathbf{H}_{0ykm}, \mathbf{H}_{0\omega kl}, \sum_{0k}$, and $\mathbf{R}_{0k}$ are hyperparameters with values that are prespecified.

Notably, we assign the conjugate prior distributions in Equation 5 because they induce the same forms of posterior distributions in the proposed model framework.

The resulting posterior distributions become easily manageable, which, in turn, leads to simple and efficient posterior sampling. In specifying the hyperparameters in Equation 5, the prespecified values represent the available prior knowledge. In general, if we have good prior information about a parameter, then we specify the hyperparameters such that the corresponding prior distribution has a mean close to that from the prior knowledge and a small variance; otherwise we specify an ad hoc mean and a large variance to make the prior noninformative. In real analysis where the true values are unknown, we might conduct a preliminary analysis to get some rough idea about hyperparameters. For example, in the osteoporosis study presented later, we can conduct a preliminary analysis using a conventional SEM with one component to obtain $\hat{\mathbf{\Lambda}}_m$ and $\hat{\mathbf{\Lambda}}_{\omega l}$ and then specify $\mathbf{\Lambda}_{0km} = \hat{\mathbf{\Lambda}}_m$ and $\mathbf{\Lambda}_{0\omega kl} = \hat{\mathbf{\Lambda}}_{\omega l}$ for $k = 1, \cdots, K$. A detailed discussion can be found in Song and Lee (2012).

## Posterior Sampling

We introduce a latent allocation variable $\mathrm{w}_i$ for $y_i$ and assume that $\mathrm{w}_i$ is independently drawn from the multinomial distribution as follows:

$$p(\mathrm{w}_i = k) = \pi_k, k = 1, \cdots, K.$$

Let $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ be the observed data matrix, $\Omega = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ be the matrix of latent variables, $\mathbf{W} = (\mathrm{w}_1, \cdots, \mathrm{w}_n)$ be the vector of latent allocation variables, and $\boldsymbol{\theta}$ be the vector of unknown parameters. A main task for the posterior inference is to sample from $p(K, \mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\Omega}|\mathbf{Y})$. However, this joint posterior distribution is intractable owing to the involvement of high-dimensional integral with respect to $K$, $\mathbf{W}$, and $\boldsymbol{\Omega}$. Thus, we employ the MCMC method, such as the Gibbs sampler and the MH algorithm, to iteratively sample from the following full conditional distributions of the unknowns.

The full conditional distribution $p(\mathbf{W}|\mathbf{Y}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(\mathrm{w}_i = k|\boldsymbol{y}_i, \boldsymbol{\theta})$ has the form of

$$p(\mathrm{w}_i = k|\boldsymbol{y}_i, \boldsymbol{\theta}) = \frac{\pi_k f_k(\boldsymbol{y}_i|\boldsymbol{\mu}_k, \sum_k)}{\sum_{k=1}^{K} \pi_k f_k(\boldsymbol{y}_i|\boldsymbol{\mu}_k, \sum_k)},$$

where $f_k(\boldsymbol{y}_i|\boldsymbol{\mu}_k, \sum_k)$ is the probability density function of $N(\boldsymbol{\mu}_k, \sum_k)$. Based on the SEM defined by Equations 2 and 3, $\sum_k = \mathbf{\Lambda}_k \sum_{\omega k} \mathbf{\Lambda}_k^T + \mathbf{\Psi}_k$, where

$$\sum_{\omega k} = \begin{bmatrix} (\boldsymbol{I} - \mathbf{\Pi}_k)^{-1}(\mathbf{\Gamma}_k \mathbf{\Phi}_k \mathbf{\Gamma}_k^T + \mathbf{\Psi}_{\delta k})(\boldsymbol{I} - \mathbf{\Pi}_k)^{-T} & (\boldsymbol{I} - \mathbf{\Pi}_k)^{-1}\mathbf{\Gamma}_k \mathbf{\Phi}_k \\ \mathbf{\Phi}_k \mathbf{\Gamma}_k^T (\boldsymbol{I} - \mathbf{\Pi}_k)^{-T} & \mathbf{\Phi}_k \end{bmatrix}.$$

The full conditional distribution $p(\boldsymbol{\Omega}|\mathbf{Y}, \mathbf{W}, \boldsymbol{\theta})$:

$$\begin{aligned} p(\boldsymbol{\Omega}|\boldsymbol{Y}, \mathbf{W}, \boldsymbol{\theta}) &= \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \mathrm{w}_i = k, \theta) \\ &\propto \prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\omega}_i, \mathrm{w}_i = k, \boldsymbol{\mu}, \theta_y) \\ &\quad p(\boldsymbol{\omega}_i|\mathrm{w}_i = k, \theta_\omega), \end{aligned}$$

and

$$p(\boldsymbol{\omega}_i|\boldsymbol{y}_i, \mathrm{w}_i = k, \boldsymbol{\theta}) \sim N(\boldsymbol{\Omega}_k^{-1} \boldsymbol{\Lambda}_k^T \boldsymbol{\Psi}_k^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_k), \boldsymbol{\Omega}_k^{-1}),$$

where $\boldsymbol{\Omega}_k = \sum_{\omega k}^{-1} + \boldsymbol{\Lambda}_k^T \boldsymbol{\Psi}_k^{-1} \boldsymbol{\Lambda}_k$.

The full conditional distribution $p(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Omega}, \mathbf{W})$: Based on $p(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Omega}, \mathbf{W}) \propto \prod_{i=1}^{n} [p(\mathrm{w}_i = k|\boldsymbol{\pi}) p(\boldsymbol{\pi})] [p(\boldsymbol{y}_i|\boldsymbol{\omega}_i, \mathrm{w}_i = k, \boldsymbol{\tau}_k) p(\boldsymbol{\tau}_k)]$, the full conditional distributions of $\boldsymbol{\pi}$ and $\boldsymbol{\tau}_k$ can be considered separately.

The full conditional for $\boldsymbol{\pi}$ is

$$p(\boldsymbol{\pi}|\mathbf{W}) \sim D(\alpha + n_1, \cdots, \alpha + n_K),$$

where $n_k$ is the sample size of the $k$th component.

Let $\mathbf{Y}_k$, $\boldsymbol{\Omega}_k$, and $\boldsymbol{\Omega}_{\xi k}$ be the submatrices of $\mathbf{Y}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Omega}_\xi = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n)$, respectively, with all the $i$ th columns with $\mathrm{w}_i = k$ being deleted. Then, we have

$$[\boldsymbol{\mu}_k|\boldsymbol{Y}_k, \boldsymbol{\Omega}_k, \boldsymbol{\Lambda}_k, \boldsymbol{\Psi}_k] \overset{D}{=} N(\boldsymbol{b}_{\mu km}, \mathbf{B}_{\mu km}),$$

$$[\mathbf{\Lambda}_{km}|\mathbf{Y}_k, \boldsymbol{\Omega}_k, \mu_{km}, \psi_{km}] \overset{D}{=} N(\boldsymbol{b}_{ykm}, \mathbf{B}_{ykm}),$$

$$[\psi_{km}^{-1}|\mathbf{Y}_k, \boldsymbol{\Omega}_\mathbf{k}, \mu_{km}] \overset{D}{=} Gamma(n_k/2 + \alpha_{0\varepsilon m}, \beta_{\varepsilon km}),$$

$$[\mathbf{\Lambda}_{wkl}|\mathbf{Y}_k, \boldsymbol{\Omega}_\mathbf{k}, \psi_{\delta kl}] \overset{D}{=} N(\boldsymbol{b}_{\omega kl}, \mathbf{B}_{\omega km}),$$

$$[\psi_{\delta kl}^{-1}|\mathbf{Y}_k, \boldsymbol{\Omega}_k] \overset{D}{=} Gamma(n_k/2 + \alpha_{0\delta k}, \beta_{\delta kl}),$$

$$[\mathbf{\Phi}_k|\boldsymbol{\Omega}_{\xi k}] \overset{D}{=} IW_{q_2}(\boldsymbol{\Omega}_{\xi k} \boldsymbol{\Omega}_{\xi k}^T + \mathbf{R}_0^{-1}, n_k + \rho_0),$$

where $\boldsymbol{b}_{\mu mk} = \mathbf{B}_{\mu mk}(n_k \boldsymbol{\Psi}_k^{-1} \bar{\mathrm{A}}_k + \sum_0^{-1} \boldsymbol{\mu}_0)$, $\mathbf{B}_{\mu mk} = (\sum_0^{-1} + n_k \mathbf{\Psi}_k^{-1})^{-1}$, $\bar{\mathbf{A}}_k = \sum_{i:\mathrm{w}_i = k} (\boldsymbol{y}_i - \boldsymbol{\Lambda}_k \boldsymbol{\xi}_i)/n_k$, and $\sum_{i:\mathrm{w}_i = k}$ denotes summation with respect to those $i$ such that $\mathrm{w}_i = k$; $\boldsymbol{b}_{ykm} = \mathbf{B}_{ykm}(\mathbf{H}_{0ykm}^{-1} \boldsymbol{\Lambda}_{0km} + \boldsymbol{\Omega}_k Y_{km}^T)$, $\mathbf{B}_{ykm} = (\mathbf{H}_{0ykm}^{-1} + \boldsymbol{\Omega}_k \boldsymbol{\Omega}_k^T)^{-1}$; $\beta_{\varepsilon km} = \beta_{0\varepsilon m} + 2^{-1}[\boldsymbol{Y}_{km} \boldsymbol{Y}_{km}^T - \omega_{ykm}^T \mathbf{B}_{ykm}^{-1} \omega_{ykm} + \mathbf{\Lambda}_{0km}^T \mathbf{H}_{0ykm}^{-1} \boldsymbol{\Lambda}_{0km}]$; $\omega_{\xi kl} = \mathbf{B}_{\xi kl}(\mathbf{H}_{0\xi kl}^{-1} \boldsymbol{\Pi}_{0kl} + \boldsymbol{\Omega}_k \boldsymbol{\eta}_{kl}^T)$, $\mathbf{B}_{\xi kl} = (\mathbf{H}_{0\xi km}^{-1} + \boldsymbol{\Omega}_k \boldsymbol{\Omega}_k^T)^{-1}$; $\beta_{\delta kl} = \beta_{0\delta k} + 2^{-1}[\boldsymbol{\eta}_{kl} \boldsymbol{\eta}_{kl}^T - \omega_{\xi kl}^T \mathbf{B}_{\xi kl}^{-1} \omega_{\xi kl} + \boldsymbol{\Pi}_{0kl}^T \mathbf{H}_{0\xi kl}^{-1} \boldsymbol{\Pi}_{0kl}]$, in which $\boldsymbol{Y}_{km}$ is the $m$th row of $\boldsymbol{Y}_k$, which is a matrix with columns that are equal to the columns of $\boldsymbol{Y}_k$ minus $\boldsymbol{\mu}_k$, and $\boldsymbol{\eta}_{kl}$ is the $l$th row of $\boldsymbol{\eta}_k$.

## Implementation of RJMCMC for Mixture Structural Equation Models

The RJMCMC method for mixture SEMs with latent variables typically contains the following key steps:

1. Update the vector of latent allocation variables $\mathbf{W}$.
2. Update parameter vector $\boldsymbol{\theta}$.
3. Update the matrix of latent variables $\boldsymbol{\Omega}$.
4. Split one component into two, or merge two components into one.

Steps 1 to 3 are conventional MCMC steps, and Step 4 is the reversible jump step, in which we allow the component number to vary from 1 to a prespecified large integer $k_{max}$ (Robert et al., 2000) and then update $\boldsymbol{\theta}$ accordingly. In Step 4, we make a random choice between attempting to split or combine, with probabilities $p_k$ and $q_k = 1 - p_k$, respectively, depending on $k$. Here, we assign $q_1 = 0$, $p_{k_{max}} = 0$, and $p_k = q_k = 0.5$, for $k = 2, 3, \cdots, k_{max} - 1$ (Boys & Henderson, 2001; Papastamoulis & Iliopoulos, 2009; Richardson & Green, 1997; Robert et al., 2000).

Richardson and Green (1997) adopted a birth-and-death step in an attempt to improve the mixing of their split/merge-based reversible jump algorithm. However, several existing works argued that this step is unnecessary in the implementation of RJMCMC. For example, Robert et al. (2000, p. 69) pointed out that "the birth or death move was completely redundant in the sense that its acceptance rates were virtually zero (below $3 * 10^{-5}$). Hence, these moves can just as well be removed from the algorithm." Boys and Henderson (2001) also discussed nonnecessity of the birth-and-death step and pointed out that the algorithm worked well without this step. Thus, we exclude the birth-and-death step in the proposed RJMCMC algorithm.

In implementing Step 4, we randomly select a component $k^*$ for split or two adjacent components $(k_1, k_2)$ for merge. Suppose that the current number of components is $K$. We generate a random number $c$ from $U(0,1)$. In a split step, we set $k^* = 1$ if $0 \le c < \frac{1}{K}$, $k^* = 2$ if $\frac{1}{K} \le c < \frac{2}{K}$, $\cdots$, $k^* = K$ if $\frac{K-1}{K} \le c \le 1$. In a merge step, we set $(k_1, k_2) = (1, 2)$ if $0 \le c < \frac{1}{K-1}$, $(k_1, k_2) = (2, 3)$ if $\frac{1}{K-1} \le c < \frac{2}{K-1}$, $\cdots$, $(k_1, k_2) = (K-1, K)$ if $\frac{K-2}{K-1} \le c \le 1$.

Next, we propose the following scheme to carry out the split/merge step. We first describe how to split component $k^*$ into components $k_1$ and $k_2$, where $k^*$ denotes the allocation of the selected component for splitting. Let $(\pi_{k^*}, \boldsymbol{\tau}_{k^*})$, $(\pi_{k_1}, \boldsymbol{\tau}_{k_1})$, and $(\pi_{k_2}, \boldsymbol{\tau}_{k_2})$ denote the probabilities and parameter vectors corresponding to components $k^*$, $k_1$, and $k_2$, respectively. In the context of mixture SEMs, the parameters in the $k$th component are $(\pi_k, \boldsymbol{\tau}_k)$, where $\boldsymbol{\tau}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k, \boldsymbol{\Lambda}_{\omega k}, \boldsymbol{\Psi}_k, \boldsymbol{\Psi}_{\delta k}, \boldsymbol{\Phi}_k\}$. The split comprises the following substeps:

1. Split the $k^*$th to $k_1$th and $k_2$th components, such that $\pi_{k^*} = \pi_{k_1} + \pi_{k_2}$. This step is accomplished by generating $a$ from $U(0,1)$ and letting $\pi_{k_1} = a\pi_{k^*}$ and $\pi_{k_2} = (1-a)\pi_{k^*}$.
2. Generate random variables $d_1, d_2$, and $d_3$ from $N(0,1)$, and $d_4$ and $d_5$ from $U(-0.5, 0.5)$, and use the random variables $d_1, \cdots, d_5$ to split $\boldsymbol{\tau}_{k^*}$ to $\boldsymbol{\tau}_{k_1}$ and $\boldsymbol{\tau}_{k_2}$ as follows:

$$
\begin{aligned}
\mu_{k_1} &= \mu_{k^*} - d_1, \quad \mu_{k_2} = \mu_{k^*} + d_1, \\
\lambda_{k_1} &= \lambda_{k^*} - d_2, \quad \lambda_{k_2} = \lambda_{k^*} + d_2, \\
\lambda_{\omega k_1} &= \lambda_{\omega k^*} - d_3, \quad \lambda_{\omega k_2} = \lambda_{\omega k^*} + d_3, \\
\psi_{k_1} &= (1 - d_4)\psi_{k^*}, \quad \psi_{k_2} = (1 + d_4)\psi_{k^*}, \\
\psi_{\delta k_1} &= (1 - d_5)\psi_{\delta k^*}, \quad \psi_{\delta k_2} = (1 + d_5)\psi_{\delta k^*},
\end{aligned} \tag{6}
$$

where $\mu_s, \lambda_s, \lambda_{\omega s}$, $\psi_s$, and $\psi_{\delta s}$ denote an arbitrary unknown element of $\boldsymbol{\mu}_s$, $\boldsymbol{\Lambda}_s$, $\boldsymbol{\Lambda}_{\omega s}$, $\boldsymbol{\Psi}_s$, and $\boldsymbol{\Psi}_{\delta s}$, respectively, for $s = k^*, k_1, k_2$. Notably, when splitting $\boldsymbol{\Phi}_k$, we must guarantee $\boldsymbol{\Phi}_k$ to be positive definite. Here, $\boldsymbol{\Phi}_k$ is a $q_2$-dimensional symmetric matrix, and thus includes $q^* = \frac{q_2(q_2+1)}{2}$ unknown elements. We generate $q^*$ random numbers $d_1, \cdots, d_{q^*}$ from $N(0,1)$ and array them in row-wise to form a $q_2$-dimensional lower triangular matrix $\boldsymbol{L}$. According to the Cholesky decomposition principle, $\boldsymbol{L}\boldsymbol{L}^T$ is a $q_2$-dimensional positive definite matrix. Thus, $\boldsymbol{\Phi}_k$ can be split (or merged) as follows:

$$
\boldsymbol{\Phi}_{k_1} = \boldsymbol{\Phi}_{k^*}, \quad \boldsymbol{\Phi}_{k_2} = \boldsymbol{L}\boldsymbol{L}^T
$$

- Accept this split step with the probabsility $min\,(1, A)$ $A = (likelihood\ ratio) \times (prior\ ratio) \times (jump\ ratio) \times (proposal\ ratio) \times |Jacobian|$, each term of which can be computed as follows:

- Likelihood ratio:

$$
\prod_{i=1}^{n} \exp\left[\frac{1}{2}\sum_{j=1}^{p}\left\{\log(\psi_{w_i,j}) + (y_{i,j} - \mu_{w_i,j} - \lambda_{w_i,j}^T\omega_i)^2/\psi_{w_i,j} - \log(\psi_{w'_i,j}) - (y_{i,j} - \mu_{w'_i,j} - \lambda_{w'_i,j}^T\omega_i)^2/\psi_{w'_i,j}\right\}\right].
$$

- Prior ratio:

$$
prior\ ratio\ of\ \boldsymbol{\pi} = \frac{\boldsymbol{\Gamma}(K+1)}{\boldsymbol{\Gamma}(K)},
$$

$$
prior\ ratio\ of\ \boldsymbol{\mu} = \frac{1}{8\pi^3|\boldsymbol{\Sigma}_0|^{0.5}}\exp\left[-\frac{1}{2}\left\{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_0) - (\boldsymbol{\mu}_{k^*} - \boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}_{k^*} - \boldsymbol{\mu}_0)\right\}\right],
$$

$$
prior\ ratio\ of\ \lambda_i = \frac{1}{\sqrt{2\pi\psi_{k^*,i}}}\exp\left[\frac{1}{2\psi_{k^*,i}}\left\{(\lambda_{k^*i} - \lambda_{0k^*i})^2 - (\lambda_{k_1 i} - \lambda_{0k^*i})^2 - (\lambda_{k_2 i} - \lambda_{0k^*i})^2\right\}\right],
$$

$$prior\ ratio\ of\ \psi_i = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(\frac{\psi_{k*i}}{\psi_{k_1 i}\psi_{k_2 i}}\right)^{\alpha-1} \exp\left\{\beta\left(\frac{1}{\psi_{k*i}}-\frac{1}{\psi_{k_1 i}}-\frac{1}{\psi_{k_2 i}}\right)\right\},$$

$$prior\ ratio\ of\ \gamma_i = \frac{1}{\sqrt{2\pi\psi_{k*,i}}}\exp\left[\frac{1}{2\psi_{k*,i}}\{(\gamma_{k*i}-\gamma_{0k*i})^2 - (\gamma_{k_1 i}-\gamma_{0k*i})^2 - (\gamma_{k_2 i}-\gamma_{0k*i})^2\}\right]$$

$$prior\ ratio\ of\ \psi_{\delta i} = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(\frac{\psi_{\delta k*i}}{\psi_{\delta k_1 i}\psi_{\delta k_2 i}}\right)^{\alpha-1} \exp\left\{\beta\left(\frac{1}{\psi_{\delta k*i}}-\frac{1}{\psi_{\delta k_1 i}}-\frac{1}{\psi_{\delta k_2 i}}\right)\right\},$$

$$prior\ ratio\ of\ \ \mathbf{\Phi} = \left\{2^{\frac{\rho_0 q_2}{2}}\pi^{\frac{q_2(q_2-1)}{4}}\prod_i^{q_2}\Gamma\left(\frac{\rho_0+1-i}{2}\right)\right\}^{-1} \frac{|\ \mathbf{\Phi}_{k_1}|^{-\frac{\rho_0+q_2+1}{2}}\exp(-\frac{1}{2}tr(\ \mathbf{\Phi}_{k_1}^{-1}))|\ \mathbf{\Phi}_{k_2}|^{-\frac{\rho_0+q_2+1}{2}}\exp(-\frac{1}{2}tr(\mathbf{\Phi}_{k_2}^{-1}))}{|\ \mathbf{\Phi}_{k*}|^{-\frac{\rho_0+q_2+1}{2}}\exp(-\frac{1}{2}tr(\ \mathbf{\Phi}_{k*}^{-1}))}.$$

- Jump ratio: $\frac{q_{k+1}}{p_k}$.

- Proposal ratio:

$$\left\{\prod_{i=1}^{p}\varphi(d_{1,i})\prod_{j=1}^{p-q}\varphi(d_{2,j})\prod_{m=1}^{q_1}\varphi(d_{3,m})\prod_{l=1}^{q_2(q_2+1)/2}\varphi(z_l)\right\}^{-1},$$

where $\varphi(\cdot)$ is the probability density function of $N(0,1)$, $d_{1,i}, d_{2,j}, d_{3,m}$ are defined similarly as $d_1, d_2, d_3$ in Equation 6, and $z_l$ is the element in lower triangle matrix $L$.

- Jacobian:

$$2^{3p+h}\pi_{k*}\prod_{i=1}^{p}\psi_{k^*i}\prod_{j=1}^{q_2}L_{jj}^{q_2+1-j}\prod_{m=1}^{q_1}\psi_{\delta k^*m},$$

where $h$ denotes the number of free elements in $\mathbf{\Lambda}_{\omega k}$.

As shown by Richardson and Green (1997) and Boys and Henderson (2001), the split in this manner ensures that (a) the parameters can jump all over the parameter space, (b) the elements of $\mathbf{\Psi}_k$ and $\mathbf{\Psi}_{\delta k}$ are positive, and (c) all the parameters can jump reversibly. After splitting one-component parameters into two, we must check whether $\mu_{k,m} < \mu_{k+1,m}$, where $k = 1, 2, \cdots, K-1$. If $\mu_{k,m} < \mu_{k+1,m}$ is not satisfied, then splitting is rejected forthwith because split and merge cannot be reversible.

The merge step can be implemented in a similar manner. We must choose a pair of adjacent components $(k_1, k_2)$ at random to ensure the split and merge can be reversible. The acceptance probability of the split move is $min(1, A)$ with $A = (likelihood\ ratio) \times (prior\ ratio) \times (jump\ ratio) \times (proposal\ ratio) \times |Jacobian|$; as such, the acceptance probability of the corresponding merge move is $min(1, A^{-1})$. In merging two adjacent components $k_1$ and $k_2$ into component $k^*$, the corresponding numbers $d_1, d_2, d_3, d_4$, and $d_5$ are generated in a reversible sense as follows:

$$\begin{aligned}
\mu_{k^*} &= \frac{\mu_{k_1}+\mu_{k_2}}{2}, \quad d_1 = \frac{\mu_{k_2}-\mu_{k_1}}{2},\\
\lambda_{k^*} &= \frac{\lambda_{k_1}+\lambda_{k_2}}{2}, \quad d_2 = \frac{\lambda_{k_2}-\lambda_{k_1}}{2},\\
\lambda_{\omega k^*} &= \frac{\lambda_{\omega k_1}+\lambda_{\omega k_2}}{2}, \quad d_3 = \frac{\lambda_{\omega k_2}-\lambda_{\omega k_1}}{2},\\
\psi_{k^*} &= \frac{\psi_{k_1}+\psi_{k_2}}{2}, \quad d_4 = \frac{\psi_{k_2}-\psi_{k_1}}{\psi_{k_1}+\psi_{k_2}},\\
\psi_{\delta k^*} &= \frac{\psi_{\delta k_1}+\psi_{\delta k_2}}{2}, \quad d_5 = \frac{\psi_{\delta k_2}-\psi_{\delta k_1}}{\psi_{\delta k_1}+\psi_{\delta k_2}}.
\end{aligned} \tag{7}$$

The merge step in Equation 7 results in $\mu_{k^*} - d_1 = \mu_{k_1}$, $\mu_{k^*} + d_1 = \mu_{k_2}$, $\lambda_{k^*} - d_2 = \lambda_{k_1}$, $\lambda_{k^*} + d_2 = \lambda_{k_2}$, $\lambda_{\omega k^*} - d_3 = \lambda_{\omega k_1}$, $\lambda_{\omega k^*} + d_3 = \lambda_{\omega k_2}$, $(1-d_4)\psi_{k^*} = \psi_{k_1}$, $(1+d_4)\psi_{k^*} = (1+d_4)\psi_{k^*} = \psi_{k_2}$, $(1-d_5)\psi_{\delta k^*} = \psi_{\delta k_1}$, and $(1+d_5)\psi_{\delta k^*} = \psi_{\delta k_2}$, which are identical to those given by Equation 6. To merge $\mathbf{\Phi}_{k_1}$ and $\mathbf{\Phi}_{k_2}$ into $\mathbf{\Phi}_{k^*}$, we set $\mathbf{\Phi}_{k^*} = \mathbf{\Phi}_{k_1}$ and conduct the Cholesky decomposition on $\mathbf{\Phi}_{k_2}$ such as $\mathbf{\Phi}_{k_2} = \boldsymbol{L}\boldsymbol{L}^T$. Then, $\boldsymbol{L}$ is a $q_2$-dimensional lower triangular matrix. The corresponding numbers $d_1, \cdots, d_{q^*}$, which are involved in computing the acceptance probability of the merge step, are taken as the elements of $\boldsymbol{L}$ in row-wise. Such a split/merge scheme guarantees that all the covariance matrices involved are positive definite and the split/merge steps are reversible.

## SIMULATION STUDY

In this section, we demonstrate the effectiveness of the proposed methodology in correctly identifying the number of components in mixture SEMs. We first consider a two-component mixture of SEMs, and then consider a four-component one.

### Simulation 1

We consider a two-component mixture of SEMs to mimic the scenario of the real data analysis. The component-specific structural equation model is defined by Equations 2 and 3 with $p = 17$, $q_1 = 2$, and $q_2 = 4$. Here, $\boldsymbol{y}_i = (y_1, \cdots, y_{17})^T$, $\boldsymbol{\omega}_i = (\eta_1, \eta_2, \xi_1, \xi_2, \xi_3, \xi_4)^T$. For notation simplicity, we temporarily omit the subscript $k$ in the component-specific parameters and specify the structures of $\mathbf{\Lambda}$, $\mathbf{\Pi}$, and $\mathbf{\Gamma}$ as follows:

$$\mathbf{\Lambda}^T = \begin{bmatrix} 1^* & \lambda_1 & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 1^* & \lambda_2 & \lambda_3 & \lambda_4 & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_5 & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_6 & \lambda_7 & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_8 & \lambda_9 & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_{10} & \lambda_{11} \end{bmatrix}, \tag{8}$$

$$\mathbf{\Pi} = \begin{bmatrix} 0^* & b \\ 0^* & 0^* \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \\ 0^* & 0^* & \gamma_5 & 0^* \end{bmatrix}, \tag{9}$$

where the elements with asterisks are fixed parameters in the model. With the prefixed parameters in $\mathbf{\Lambda}$, the factor loading matrix has a nonoverlapping structure, which implies that the observed variables $\{y_1, y_2\}$, $\{y_3, y_4, y_5, y_6\}$, $\{y_7, y_8\}$, $\{y_9, y_{10}, y_{11}\}$, $\{y_{12}, y_{13}, y_{14}\}$, and $\{y_{15}, y_{16}, y_{17}\}$ characterize the latent variables $\eta_1, \eta_2$, $\xi_1$, $\xi_2$, $\xi_3$, and $\xi_4$, respectively, and no observed variable characterizes more than one latent variable. One of the elements at each column of $\mathbf{\Lambda}$ is fixed at 1 to obtain an identified model and unify the scales of the latent variables.

The true population values of the unknown parameters are set as follows:

Component I: $\pi_1 = 0.5$, $\mu_1 = \mu_2 = \cdots = \mu_{17} = 0$, $\lambda_1 = \lambda_2 = \cdots = \lambda_{11} = -1.5$, $\psi_1 = \psi_2 = \cdots = \psi_{17} = 0.5$, $b = 0.5$, $\gamma_1 = \gamma_2 = \cdots = \gamma_5 = 0.5$, $\psi_{\delta 1} = \psi_{\delta 2} = 0.6$, $\{\phi_{11}, \phi_{12}, \phi_{13}, \phi_{14}, \phi_{22}, \phi_{23}, \phi_{24}, \phi_{33}, \phi_{34}, \phi_{44}\} = \{1, 0.1, 0, 0, 1, 0.2, 0.3, 1, 0.8, 1\}$.

Component II: $\pi_2 = 0.5$, $\mu_1 = \mu_2 = \cdots = \mu_{17} = 2$, $\lambda_1 = \lambda_2 = \cdots = \lambda_{11} = 2$, $\psi_1 = \psi_2 = \cdots = \psi_{17} = 0.6$, $b = 0.3$, $\gamma_1 = \gamma_2 = \cdots = \gamma_5 = 1.5$, $\psi_{\delta 1} = \psi_{\delta 2} = 0.7$, $\{\phi_{11}, \phi_{12}, \phi_{13}, \phi_{14}, \phi_{22}, \phi_{23}, \phi_{24}, \phi_{33}, \phi_{34}, \phi_{44}\} = \{1, 0.2, 0, 0, 1, 0.3, 0.5, 1, 0.9, 1\}$.

On the basis of this setting, we generated 1,400 observations, of which approximately half come from the first component and half come from the second component. The hyperparameter $\alpha$ in the Dirichlet prior (Equation 4) is set as $\alpha = 1$, and the prior inputs involved in Equation 5 are set as follows:

Prior 1: For $k = 1, 2$, the elements of $\mathbf{\Lambda}_{0km}$ and $\mathbf{\Lambda}_{0\omega kl}$ are taken to be the true values of $\mathbf{\Lambda}_{km}$ and $\mathbf{\Lambda}_{\omega kl}$, respectively, $\mathbf{H}_{pykm} = \mathbf{I}$, $\mathbf{H}_{0\omega kl} = \mathbf{I}$, and $\mathbf{R}_{0k} = 8\mathbf{I}$, where $\mathbf{I}$ is the identity matrix with an appropriate dimension; $\alpha_{0\varepsilon k} = \alpha_{0\delta k} = 6$, $\beta_{0\varepsilon k} = \beta_{0\delta k} = 4$, $\rho_{0k} = 8$, $\boldsymbol{\mu}_0 = \bar{\boldsymbol{y}}$, and $\mathbf{\Sigma}_0 = \boldsymbol{S}_y/2$, where $\bar{\boldsymbol{y}}$ and $\boldsymbol{S}_y$ are the sample mean and the sample covariance matrix of the simulated data set.

Prior 2: For $k = 1, 2$, the elements of $\mathbf{\Lambda}_{0km}$ and $\mathbf{\Lambda}_{0\omega kl}$ are taken to be 0, $\mathbf{H}_{pykm} = \mathbf{I}$, $\mathbf{H}_{0\omega kl} = \mathbf{I}$, $\mathbf{R}_{0k} = 8\mathbf{I}$, $\alpha_{0\varepsilon k} = \alpha_{0\delta k} = 2$, $\beta_{0\varepsilon k} = \beta_{0\delta k} = 4$, $\rho_{0k} = 8$, $\boldsymbol{\mu}_0 = \mathbf{0}$, and $\mathbf{\Sigma}_0 = \mathbf{I}$.

Let $M_K$ be a $K$-component mixture of SEMs. In implementing the RJMCMC algorithm, we set $k_{max} = 100$, which provides sufficient flexibility in accommodating various complex situations. In each replication, we collect 5,000 posterior samples after discarding 5,000 burn-in iterations. Table 1 (second and third columns) reports the estimated posterior probability of $M_K$ summarized based on 100 replications. $M_2$ is selected with the highest estimated posterior probability under both Prior 1 and Prior 2. Figure 1 presents

TABLE 1
Posterior Probabilities of $K$-Component Models

| | Simulation 1 | | Simulation 2 | | Real Example | |
|---|---|---|---|---|---|---|
| K | Prior 1 | Prior 2 | Prior 1 | Prior 2 | Prior 1 | Prior 2 |
| 1 | 0.203 (0.038) | 0.127 (0.058) | 0.003 (0.004) | 0.004 (0.006) | 0.243 | 0.268 |
| 2 | 0.519 (0.039) | 0.428 (0.061) | 0.038 (0.027) | 0.039 (0.034) | 0.431 | 0.371 |
| 3 | 0.219 (0.041) | 0.300 (0.052) | 0.240 (0.060) | 0.241 (0.073) | 0.280 | 0.269 |
| 4 | 0.049 (0.020) | 0.107 (0.034) | 0.369 (0.043) | 0.368 (0.049) | 0.046 | 0.088 |
| 5 | 0.010 (0.008) | 0.030 (0.015) | 0.225 (0.044) | 0.221 (0.056) | 0.001 | 0.004 |
| 6 | 0.002 (0.003) | 0.007 (0.007) | 0.090 (0.039) | 0.092 (0.041) | 0.000 | 0.001 |
| 7 | 0.000 (0.001) | 0.001 (0.003) | 0.027 (0.020) | 0.028 (0.019) | | |
| 8 | 0.000 (0.001) | 0.000 (0.001) | 0.006 (0.007) | 0.007 (0.007) | | |
| 9 | | | 0.001 (0.002) | 0.001 (0.003) | | |
| 10 | | | 0.000 (0.000) | 0.000 (0.002) | | |

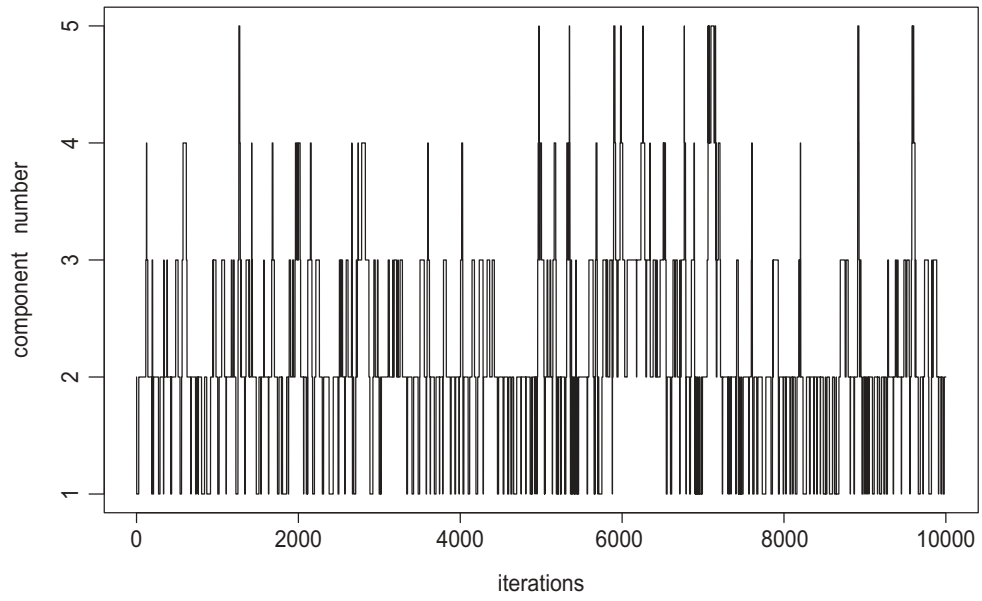



FIGURE 1 The trace plot of the number of components under Prior 1 in Simulation 1.

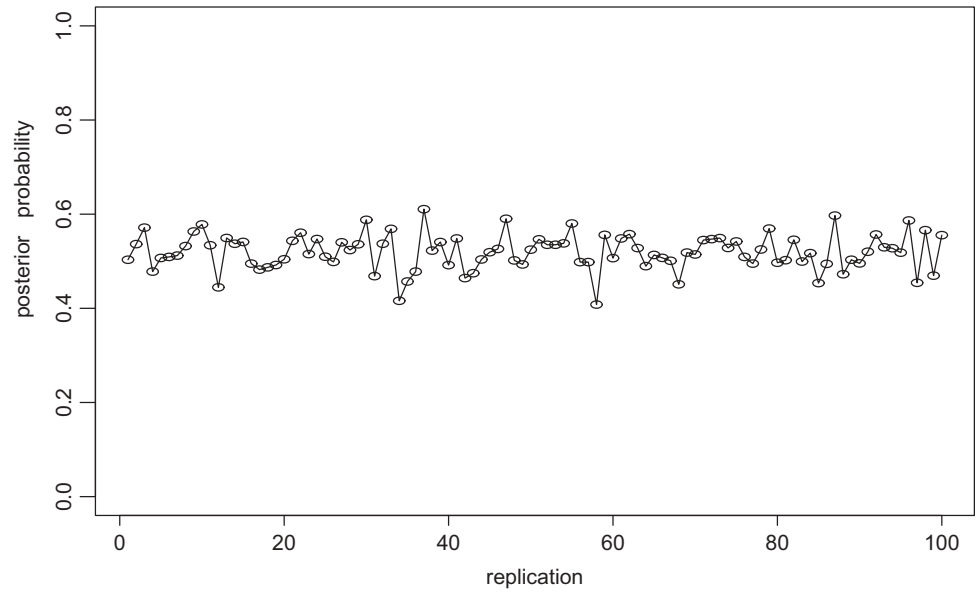



FIGURE 2 Posterior probability of $K = 2$ under Prior 1 in Simulation 1.

the trace plot of $K$ in an arbitrarily selected replication. Apparently, $K$ takes the value of 2 most frequently in the MCMC chain, and the estimated posterior probability of $M_2$ (i.e., the number of $K = 2$ among the 5,000 iterations) is among the largest. Figure 2 shows that the estimated posterior probability of $M_2$ ranges from 0.4 to 0.6 across the 100 replicated data sets with a standard deviation of 0.039. Thus, the selection result is in line with the true model setup consistently in all the replications.

In this simulation and subsequent analyses, the initial value of $K$, denoted by $K_0$, is 2. Usually, $K_0$ can be taken as any small integer. To examine how $K_0$ can affect the RJMCMC procedure, we reanalyze the previous 100 replicated data sets through the RJMCMC algorithm from different initial values of $K$. Table 2 shows that the model selection results obtained under Prior 2 and distinct values $K_0 = 1, 2, 3$ are quite similar. The proposed method performs consistently well regardless of the value of $K_0$.

For comparison, we calculate the Deviance Information Criterion (DIC) values based on the previous 100 replicated data sets. The means and standard deviations of DIC values obtained under Prior 2 for $M_k$, $K = 1, \cdots, 8$ are reported in

TABLE 2
Posterior Probabilities of $M_k$ With Different Initial Values in Simulation 1

| K | $K_0 = 1$ | $K_0 = 2$ | $K_0 = 3$ |
|---|---|---|---|
| 1 | 0.189 (0.039) | 0.203 (0.038) | 0.191 (0.042) |
| 2 | 0.514 (0.048) | 0.519 (0.039) | 0.514 (0.037) |
| 3 | 0.228 (0.042) | 0.219 (0.041) | 0.229 (0.042) |
| 4 | 0.055 (0.024) | 0.049 (0.020) | 0.052 (0.021) |
| 5 | 0.011 (0.008) | 0.010 (0.008) | 0.011 (0.008) |
| 6 | 0.002 (0.003) | 0.002 (0.003) | 0.002 (0.003) |
| 7 | 0.000 (0.001) | 0.000 (0.001) | 0.000 (0.001) |
| 8 | 0.000 (0.001) | 0.000 (0.001) | 0.000 (0.001) |

TABLE 3
Summary of the DIC Values of $K$-Component Models in Simulation 1

| $K$ | M | SD |
|---|---|---|
| 1 | 69276.624 | 444.406 |
| 2 | 53414.839 | 565.648 |
| 3 | 74981.341 | 1294.320 |
| 4 | 80581.115 | 1648.520 |
| 5 | 83383.707 | 1617.144 |
| 6 | 85329.635 | 1412.406 |
| 7 | 86683.220 | 1424.767 |
| 8 | 87454.258 | 1344.114 |

Table 3, which indicates that $M_2$ is selected with the smallest DIC. This result agrees with that obtained through the proposed RJMCMC procedure. However, the DIC-based model selection is highly computationally demanding because DIC compares models in a pairwise basis. Instead, the proposed procedure can facilitate an automatic exploration in a large model space and enable a simultaneous selection among a number of candidate models. In this simulation, the RJMCMC algorithm completes a replication with $k_{max} = 100$ every 9 min, which implies that our procedure selects the best model among 100 competing models in 9 min. By contrast, the use of DIC takes approximately 18 min to complete the pairwise comparison of $(M_1, M_2)$, $(M_2, M_3)$, and $(M_3, M_4)$. When the model space becomes large, such pairwise comparison can be very tedious and the associated computing time increases dramatically. Another popular Bayesian model selection statistic is the Bayes factor. However, this approach compares models also in a pairwise basis. The computation of Bayes factor is even more challenging and time-consuming than that of DIC. Given the high computational efficiency of the RJMCMC algorithm, our procedure is particularly attractive in the case where model selection is expected to be conducted in a large model space.

We also calculate the bias and the root mean square (RMS) between parameter estimates and their true population values to assess the empirical performance of Bayesian estimation. Tables 4 and 5 summarize the estimation results of $M_2$ under Priors 1 and 2 on the basis of 100 replications. Most of the parameter estimates perform similarly and satisfactorily with small values of bias and RMS under both prior settings. Thus, the Bayesian model selection and estimation results are not very sensitive to the prior specifications under consideration.

TABLE 4
Parameter Estimates Under $M_2$ and Prior 1 in Simulation I

| | Component 1 | | Component 2 | | | Component 1 | | Component 2 | |
|---|---|---|---|---|---|---|---|---|---|
| Par. | Bias | RMS | Bias | RMS | Par. | Bias | RMS | Bias | RMS |
| $\mu_1$ | 0.001 | 0.085 | −0.001 | 0.182 | $\psi_1$ | 0.052 | 0.059 | 0.000 | 0.043 |
| $\mu_2$ | −0.004 | 0.113 | −0.007 | 0.037 | $\psi_2$ | 0.012 | 0.069 | 0.029 | 0.095 |
| $\mu_3$ | 0.006 | 0.048 | 0.005 | 0.071 | $\psi_3$ | 0.006 | 0.028 | 0.003 | 0.031 |
| $\mu_4$ | −0.006 | 0.063 | 0.012 | 0.145 | $\psi_4$ | 0.004 | 0.036 | −0.002 | 0.041 |
| $\mu_5$ | −0.008 | 0.065 | 0.013 | 0.143 | $\psi_5$ | 0.007 | 0.038 | 0.007 | 0.043 |
| $\mu_6$ | −0.011 | 0.069 | 0.011 | 0.149 | $\psi_6$ | 0.007 | 0.042 | 0.014 | 0.052 |
| $\mu_7$ | 0.003 | 0.045 | −0.012 | 0.046 | $\psi_7$ | 0.002 | 0.045 | −0.001 | 0.037 |
| $\mu_8$ | 0.001 | 0.053 | −0.015 | 0.075 | $\psi_8$ | 0.019 | 0.076 | 0.022 | 0.085 |
| $\mu_9$ | 0.001 | 0.046 | 0.004 | 0.055 | $\psi_9$ | 0.008 | 0.036 | 0.010 | 0.036 |
| $\mu_{10}$ | −0.008 | 0.068 | −0.007 | 0.084 | $\psi_{10}$ | 0.010 | 0.046 | −0.003 | 0.053 |
| $\mu_{11}$ | −0.003 | 0.069 | −0.001 | 0.084 | $\psi_{11}$ | −0.000 | 0.044 | −0.000 | 0.058 |
| $\mu_{12}$ | −0.003 | 0.051 | 0.002 | 0.053 | $\psi_{12}$ | 0.008 | 0.032 | 0.006 | 0.037 |
| $\mu_{13}$ | 0.007 | 0.068 | 0.006 | 0.090 | $\psi_{13}$ | 0.009 | 0.039 | −0.001 | 0.050 |
| $\mu_{14}$ | 0.003 | 0.069 | 0.005 | 0.091 | $\psi_{14}$ | 0.002 | 0.032 | −0.005 | 0.043 |
| $\mu_{15}$ | −0.006 | 0.043 | −0.000 | 0.045 | $\psi_{15}$ | 0.002 | 0.035 | −0.004 | 0.032 |
| $\mu_{16}$ | 0.001 | 0.060 | 0.005 | 0.078 | $\psi_{16}$ | −0.002 | 0.043 | −0.006 | 0.047 |
| $\mu_{17}$ | −0.002 | 0.065 | 0.005 | 0.079 | $\psi_{17}$ | 0.008 | 0.038 | −0.013 | 0.055 |
| $\lambda_1$ | 0.065 | 0.267 | 0.002 | 0.016 | $\phi_{11}$ | 0.003 | 0.077 | 0.005 | 0.074 |
| $\lambda_2$ | −0.003 | 0.060 | 0.012 | 0.040 | $\phi_{12}$ | 0.004 | 0.046 | −0.005 | 0.038 |
| $\lambda_3$ | −0.005 | 0.055 | 0.011 | 0.036 | $\phi_{13}$ | 0.007 | 0.048 | −0.000 | 0.042 |
| $\lambda_4$ | −0.009 | 0.060 | 0.007 | 0.040 | $\phi_{14}$ | 0.007 | 0.044 | −0.002 | 0.041 |
| $\lambda_5$ | 0.001 | 0.066 | −0.005 | 0.074 | $\phi_{22}$ | −0.015 | 0.073 | −0.008 | 0.078 |
| $\lambda_6$ | −0.020 | 0.063 | 0.012 | 0.065 | $\phi_{23}$ | 0.002 | 0.042 | 0.001 | 0.036 |
| $\lambda_7$ | −0.023 | 0.059 | 0.014 | 0.070 | $\phi_{24}$ | 0.001 | 0.046 | −0.001 | 0.041 |
| $\lambda_8$ | 0.002 | 0.050 | −0.013 | 0.066 | $\phi_{33}$ | 0.010 | 0.081 | 0.022 | 0.081 |
| $\lambda_9$ | −0.001 | 0.052 | −0.023 | 0.070 | $\phi_{34}$ | 0.080 | 0.060 | 0.005 | 0.063 |
| $\lambda_{10}$ | 0.002 | 0.047 | −0.028 | 0.075 | $\phi_{44}$ | −0.010 | 0.075 | 0.033 | 0.081 |
| $\lambda_{11}$ | 0.006 | 0.052 | −0.022 | 0.069 | $\gamma_1$ | −0.001 | 0.075 | 0.002 | 0.064 |
| $\gamma_2$ | −0.000 | 0.082 | 0.015 | 0.084 | $\gamma_3$ | −0.012 | 0.105 | 0.007 | 0.183 |
| $\gamma_4$ | 0.004 | 0.099 | −0.050 | 0.153 | $\gamma_5$ | 0.005 | 0.036 | −0.022 | 0.063 |
| $\pi$ | 0.001 | 0.013 | −0.001 | 0.013 | b | −0.003 | 0.087 | 0.007 | 0.070 |
| $\psi_{\delta 1}$ | −0.007 | 0.069 | −0.052 | 0.100 | $\psi_{\delta 2}$ | −0.004 | 0.051 | −0.019 | 0.054 |

*Note*. Par. = parameter; RMS = root mean square.

## Simulation 2

In this section, we consider a four-component mixture of SEMs to accommodate relatively complex situations. The true population values of parameters are set as follows:

Component 1: $\pi_1 = 0.25$, $\mu_1 = \mu_2 = \cdots = \mu_{17} = -2$, $\lambda_1 = \lambda_2 = \cdots = \lambda_{11} = -2$, $\psi_1 = \psi_2 = \cdots = \psi_{17} = 0.8, b = 0.5$, $\gamma_1 = \gamma_2 = \cdots = \gamma_5 = 0.5$, $\psi_{\delta 1} = \psi_{\delta 2} = 0.8$, $\{\phi_{11}, \phi_{12}, \phi_{13}, \phi_{14}, \phi_{22}, \phi_{23}, \phi_{24}, \phi_{33}, \phi_{34}, \phi_{44}\} = \{1, 0.1, 0, 0, 1, 0.2, 0.3, 1, 0.8, 1\}$.

Component 2: $\pi_2 = 0.25$, $\mu_1 = \mu_2 = \cdots = \mu_{17} = 0$, $\lambda_1 = \lambda_2 = \cdots = \lambda_{11} = -1$, $\psi_1 = \psi_2 = \cdots = \psi_{17} = 0.6$, $b = 0.3$, $\gamma_1 = \gamma_2 = \cdots = \gamma_5 = 0.5$, $\psi_{\delta 1} = \psi_{\delta 2} = 0.7$, $\{\phi_{11},$

TABLE 5
Parameter Estimates Under $M_2$ and Prior 2 in Simulation I

| | Component 1 | | Component 2 | | | Component 1 | | Component 2 | |
|---|---|---|---|---|---|---|---|---|---|
| Par. | Bias | RMS | Bias | RMS | Par. | Bias | RMS | Bias | RMS |
| $\mu_1$ | −0.005 | 0.071 | 0.005 | 0.191 | $\psi_1$ | −0.009 | 0.035 | 0.001 | 0.039 |
| $\mu_2$ | 0.009 | 0.102 | 0.003 | 0.177 | $\psi_2$ | 0.021 | 0.133 | 0.007 | 0.120 |
| $\mu_3$ | −0.005 | 0.046 | 0.004 | 0.070 | $\psi_3$ | 0.003 | 0.030 | 0.006 | 0.032 |
| $\mu_4$ | 0.001 | 0.057 | −0.003 | 0.137 | $\psi_4$ | 0.037 | 0.051 | 0.009 | 0.039 |
| $\mu_5$ | 0.003 | 0.061 | 0.003 | 0.135 | $\psi_5$ | 0.034 | 0.048 | 0.010 | 0.040 |
| $\mu_6$ | 0.001 | 0.057 | 0.003 | 0.132 | $\psi_6$ | 0.038 | 0.050 | 0.011 | 0.048 |
| $\mu_7$ | 0.003 | 0.043 | −0.005 | 0.051 | $\psi_7$ | −0.039 | 0.050 | 0.006 | 0.043 |
| $\mu_8$ | −0.002 | 0.061 | −0.001 | 0.073 | $\psi_8$ | 0.017 | 0.182 | −0.008 | 0.081 |
| $\mu_9$ | −0.003 | 0.040 | −0.007 | 0.051 | $\psi_9$ | 0.001 | 0.028 | 0.005 | 0.033 |
| $\mu_{10}$ | 0.005 | 0.063 | 0.002 | 0.082 | $\psi_{10}$ | 0.037 | 0.053 | −0.009 | 0.056 |
| $\mu_{11}$ | 0.000 | 0.054 | −0.003 | 0.081 | $\psi_{11}$ | 0.058 | 0.067 | 0.016 | 0.051 |
| $\mu_{12}$ | −0.003 | 0.049 | 0.008 | 0.052 | $\psi_{12}$ | 0.009 | 0.034 | −0.005 | 0.033 |
| $\mu_{13}$ | 0.005 | 0.069 | 0.007 | 0.086 | $\psi_{13}$ | 0.046 | 0.060 | 0.010 | 0.048 |
| $\mu_{14}$ | 0.003 | 0.063 | 0.012 | 0.086 | $\psi_{14}$ | 0.047 | 0.060 | −0.008 | 0.036 |
| $\mu_{15}$ | −0.005 | 0.044 | −0.003 | 0.051 | $\psi_{15}$ | −0.034 | 0.101 | 0.009 | 0.097 |
| $\mu_{16}$ | 0.012 | 0.062 | 0.013 | 0.092 | $\psi_{16}$ | 0.024 | 0.051 | 0.008 | 0.046 |
| $\mu_{17}$ | 0.013 | 0.062 | 0.009 | 0.095 | $\psi_{17}$ | 0.011 | 0.043 | −0.003 | 0.042 |
| $\lambda_1$ | 0.010 | 0.036 | 0.001 | 0.014 | $\phi_{11}$ | 0.007 | 0.106 | 0.003 | 0.088 |
| $\lambda_2$ | 0.049 | 0.069 | 0.009 | 0.037 | $\phi_{12}$ | 0.004 | 0.043 | 0.002 | 0.038 |
| $\lambda_3$ | 0.041 | 0.066 | 0.011 | 0.038 | $\phi_{13}$ | 0.004 | 0.045 | −0.001 | 0.043 |
| $\lambda_4$ | 0.008 | 0.068 | 0.010 | 0.040 | $\phi_{14}$ | −0.006 | 0.061 | 0.002 | 0.079 |
| $\lambda_5$ | 0.026 | 0.099 | −0.005 | 0.076 | $\phi_{22}$ | 0.013 | 0.084 | 0.003 | 0.069 |
| $\lambda_6$ | −0.024 | 0.054 | 0.004 | 0.068 | $\phi_{23}$ | 0.009 | 0.048 | 0.002 | 0.043 |
| $\lambda_7$ | 0.027 | 0.058 | 0.006 | 0.061 | $\phi_{24}$ | 0.007 | 0.068 | 0.005 | 0.039 |
| $\lambda_8$ | 0.030 | 0.054 | −0.018 | 0.063 | $\phi_{33}$ | 0.003 | 0.079 | 0.023 | 0.077 |
| $\lambda_9$ | −0.009 | 0.055 | −0.015 | 0.065 | $\phi_{34}$ | −0.009 | 0.081 | 0.006 | 0.189 |
| $\lambda_{10}$ | 0.012 | 0.028 | 0.019 | 0.087 | $\phi_{44}$ | −0.006 | 0.278 | 0.008 | 0.225 |
| $\lambda_{11}$ | 0.027 | 0.027 | −0.047 | 0.083 | $\gamma_1$ | −0.007 | 0.043 | 0.009 | 0.065 |
| $\gamma_2$ | −0.008 | 0.044 | 0.006 | 0.085 | $\gamma_3$ | 0.003 | 0.071 | −0.001 | 0.178 |
| $\gamma_4$ | 0.006 | 0.091 | −0.005 | 0.060 | $\gamma_5$ | −0.009 | 0.043 | −0.007 | 0.066 |
| $\pi$ | −0.001 | 0.013 | 0.001 | 0.013 | b | 0.010 | 0.047 | −0.020 | 0.065 |
| $\psi_{\delta 1}$ | −0.001 | 0.048 | −0.002 | 0.079 | $\psi_{\delta 2}$ | 0.003 | 0.056 | 0.000 | 0.052 |

*Note*. Par. = parameter; RMS = root mean square.

$\phi_{12}, \phi_{13}, \phi_{14}, \phi_{22}, \phi_{23}, \phi_{24}, \phi_{33}, \phi_{34}, \phi_{44}\} = \{1, 0.2, 0, 0, 1, 0.3, 0.5, 1, 0.9, 1\}$.

Component 3: $\pi_3 = 0.25$, $\mu_1 = \mu_2 = \cdots = \mu_{17} = 1$, $\lambda_1 = \lambda_2 = \cdots = \lambda_{11} = 1$, $\psi_1 = \psi_2 = \cdots = \psi_{17} = 0.7$, $b = -0.5$, $\gamma_1 = \gamma_2 = \cdots = \gamma_5 = 0$, $\psi_{\delta 1} = \psi_{\delta 2} = 0.6$, $\{\phi_{11}, \phi_{12}, \phi_{13}, \phi_{14}, \phi_{22}, \phi_{23}, \phi_{24}, \phi_{33}, \phi_{34}, \phi_{44}\} = \{1, 0, 0, 0, 1, 0.3, 0.5, 1, 0.2, 1\}$.

Component 4: $\pi_4 = 0.25$, $\mu_1 = \mu_2 = \cdots = \mu_{17} = 2$, $\lambda_1 = \lambda_2 = \cdots = \lambda_{11} = 2$, $\psi_1 = \psi_2 = \cdots = \psi_{17} = 0.5$, $b = -1$, $\gamma_1 = \gamma_2 = \cdots = \gamma_5 = -1.5$, $\psi_{\delta 1} = \psi_{\delta 2} = 0.5$, $\{\phi_{11}, \phi_{12}, \phi_{13}, \phi_{14}, \phi_{22}, \phi_{23}, \phi_{24}, \phi_{33}, \phi_{34}, \phi_{44}\} = \{1, 0.8, 0, 0, 1, 0.3, 0.4, 1, 0.6, 1\}$.

We simulate 1,400 observations according to these settings. All the prior inputs are similar to those in Simulation 1. In the implementation of the RJMCMC algorithm, we collect 5,000 posterior samples after discarding 5,000 burn-in iterations. Table 1 (fourth and fifth columns) reports the estimated posterior probability of $M_K$ summarized based on 100 replications. $M_4$ is selected with the highest estimated posterior probability under both Prior 1 and Prior 2. Figure 3 presents the trace plot of $K$ in an arbitrarily selected replication. Likewise, $K$ takes the value of 4 most frequently in the MCMC chain, and the estimated posterior probability of $M_4$ is among the largest. Figure 4 shows that the estimated posterior probability of $M_4$ ranges from 0.3 to 0.5 across the 100 replicated data sets with a standard deviation of 0.043. Thus, the selection result is consistent with the true model setup. We likewise conduct parameter estimation for $M_4$ under Prior 1 and Prior 2. The parameter estimates (not reported) perform similarly and satisfactorily under both prior settings.

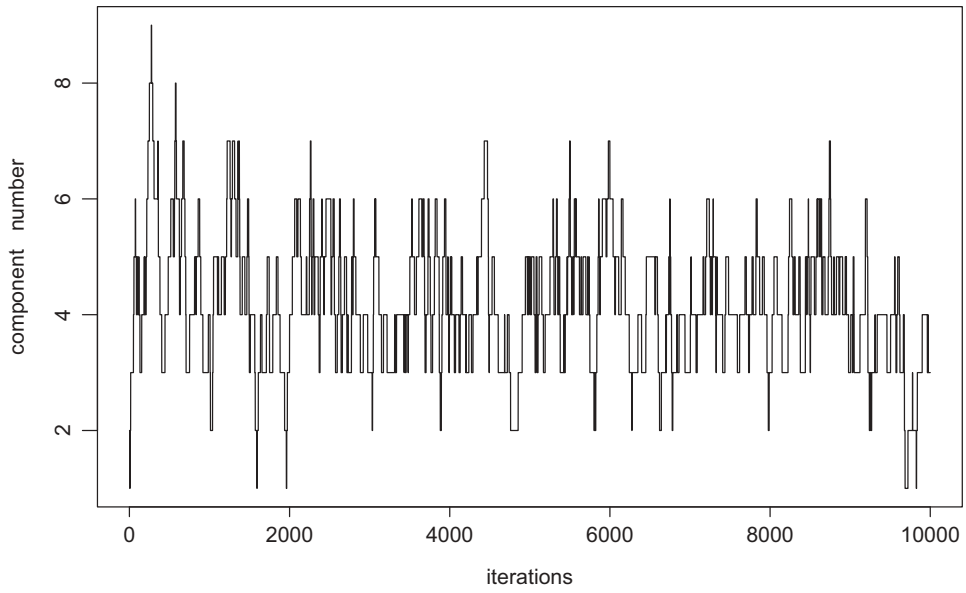



FIGURE 3 The trace plot of the number of components under Prior 1 in Simulation 2.

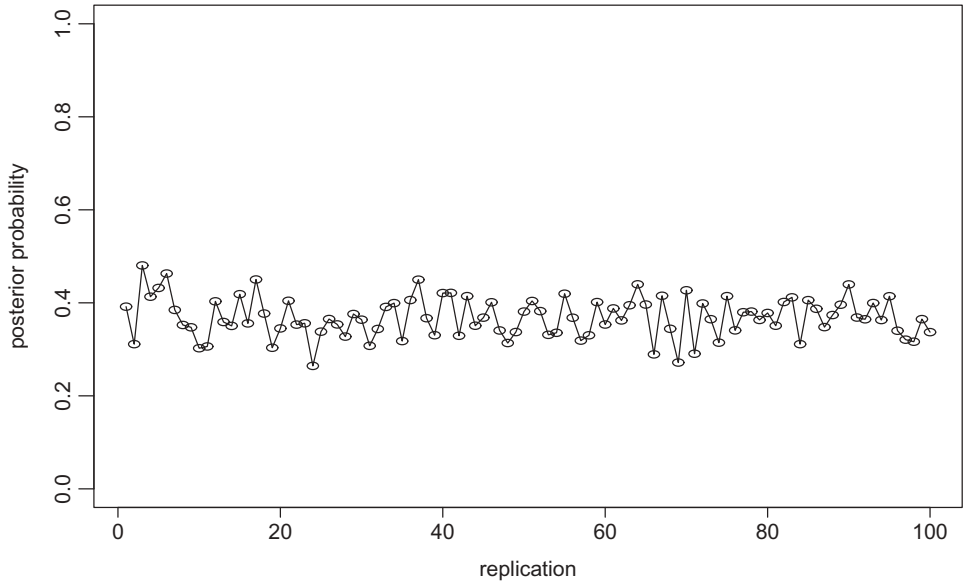



FIGURE 4 Posterior probability of $K = 4$ under Prior 1 in Simulation 2.

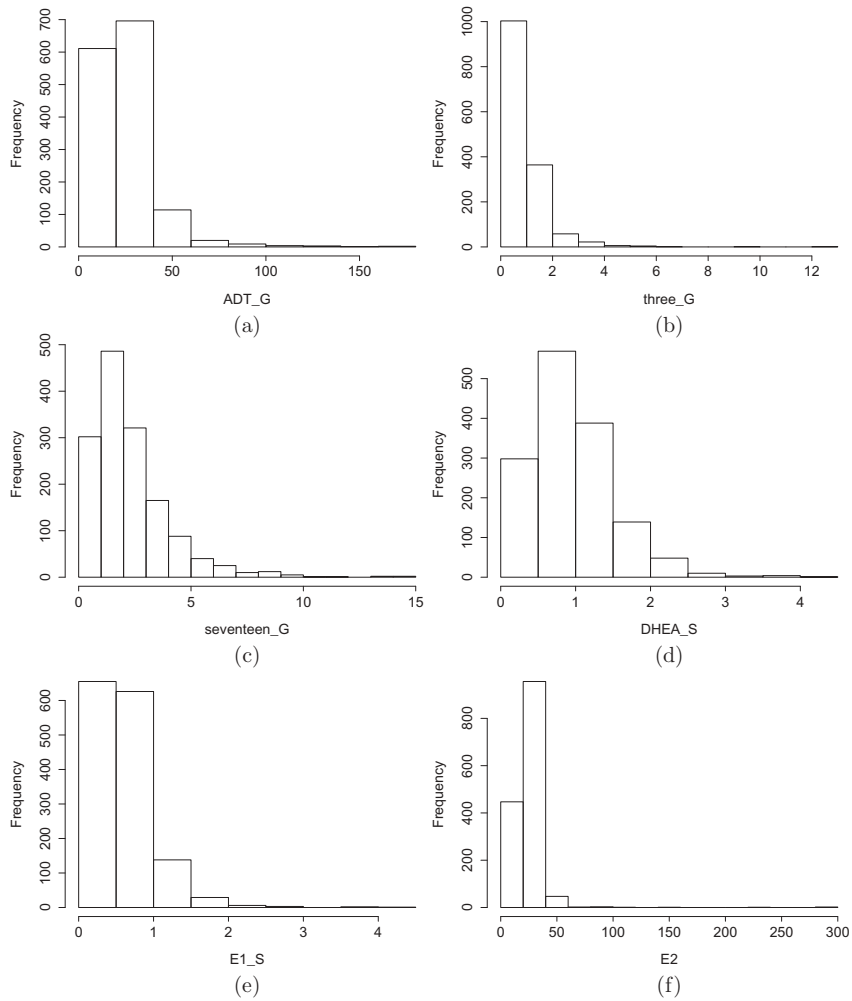



FIGURE 5 From (a) to (f), the histograms of several observed variables in the real example.

## STUDY OF THE RISK FACTORS OF OSTEOPOROTIC FRACTURES

In this section, we apply the proposed methodology to a study of the influence of serum concentrations of sex hormones, their precursors, and metabolites on bone mineral density in older men. This study was part of a multicenter prospective cohort study of risk factors of osteoporotic fractures in older people (Lau et al., 2006), and the results are important for osteoporosis prevention and control. A total of 1446 Chinese men aged 65 years and older were recruited using a combination of private solicitation and public advertising from community centers and public housing estates. The primary concern of this study is to examine the relative importance of obesity, the major sex hormones, precursors, and metabolites on BMD, so that the most appropriate form of hormonal supplementation can be devised to prevent osteoporosis for older men. Here, BMD and its important determinants, such as obesity, estrogen, androgen, precursors, and metabolites, are latent traits that are measured through $\{\text{Body mass index (BMI)}, \text{Waist hip ratio}\}$, $\{\text{spine BMD,hip BMD}\}$, $\{\text{E1}, \text{E1}-\text{S}, \text{E2}\}$, $\{\text{TESTO}, 5-\text{DIOL}, \text{DHT}\}$, $\{4-\text{DIONE}, \text{DHEA}, \text{DHEA}-\text{S}\}$, and $\{\text{ADT}, \text{ADT}-\text{G}, 3\text{G}, 3\text{G}-17\text{G}\}$, respectively. Figure 5 shows that several observed variables are far from normal, implying that heterogeneity exists in the data.

Given these features, two relevant questions are naturally raised: (a) How many mixture components exist in the data? (b) What are the component-specific effects of obesity, estrogen, androgen, precursors, and metabolites on the progression of loss of BMD? The mixture SEMs considered in this study perfectly accommodate these concerns. Based on medical literature, androgen influences metabolites, and metabolites together with precursors and the major hormones influence BMD. In addition, obesity grouped by BMI and waist–hip ratio might affect BMD. Thus, we considered BMD ($\eta_1$) and metabolites ($\eta_2$) as outcome latent variables, and obesity ($\xi_1$), estrogen ($\xi_2$), androgen ($\xi_3$), and precursors ($\xi_4$) as explanatory latent variables. The SEM was defined by Equation 2 with the factor loading matrix in Equation 8 and the following structural equations:

$$\begin{aligned} \eta_1 &= b\eta_2 + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \gamma_4\xi_4 + \delta_1, \\ \eta_2 &= \gamma_5\xi_3 + \delta_2. \end{aligned} \tag{10}$$

To implement the RJMCMC procedure, we set the prior inputs to be the same as those of Simulation 1, in which the true values of $\mathbf{\Lambda}_{km}$ and $\mathbf{\Lambda}_{\omega kl}$ in Prior 1 were replaced by the preliminary estimates of $\mathbf{\Lambda}_{km}$ and $\mathbf{\Lambda}_{\omega kl}$ obtained by conducting a conventional SEM with one component. We ran the RJMCMC algorithm for 20,000 iterations and discarded the first 10,000 iterations. Table 1 shows that the two-component mixture of SEMs was selected with the largest estimated posterior probability under Prior 1 and Prior 2. On the basis of the selected model, the Bayesian estimates of model parameters were obtained through 10,000 posterior samples after 10,000 burn-ins under both prior settings. The estimation results obtained under Prior 2 are reported in Table 6 and Figure 6. Those obtained under Prior 1 are similar and not reported.

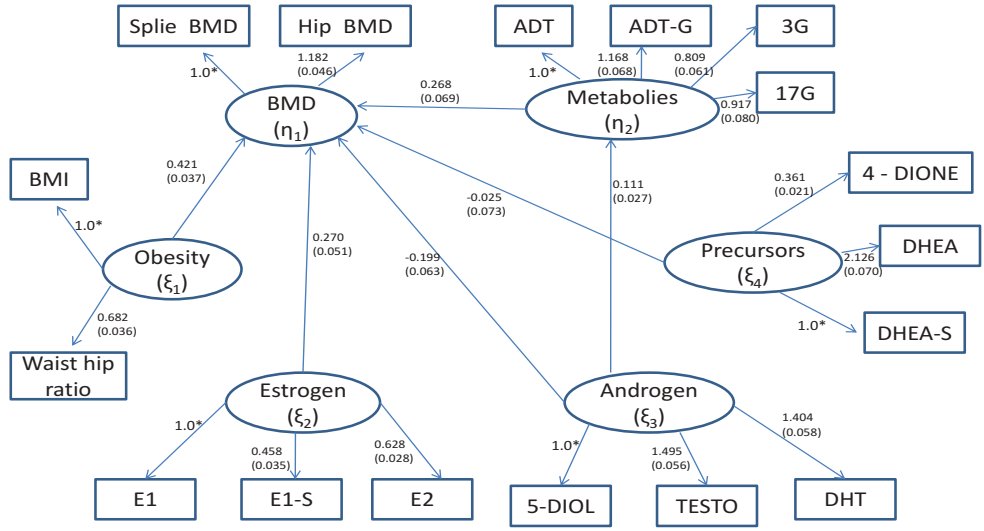
TABLE 6
Parameter Estimates Under Prior 2 in the Real Example

| | Component 1 | | Component 2 | | | Component 1 | | Component 2 | |
|---|---|---|---|---|---|---|---|---|---|
| Par. | Est | SE | Est | SE | Par. | Est | SE | Est | SE |
| $\mu_1$ | −0.058 | 0.029 | 0.352 | 0.086 | $\psi_1$ | 0.373 | 0.024 | 0.431 | 0.075 |
| $\mu_2$ | −0.043 | 0.030 | 0.257 | 0.074 | $\psi_2$ | 0.215 | 0.025 | 0.317 | 0.052 |
| $\mu_3$ | −0.111 | 0.028 | 0.601 | 0.095 | $\psi_3$ | 0.695 | 0.032 | 1.572 | 0.165 |
| $\mu_4$ | −0.186 | 0.019 | 1.070 | 0.127 | $\psi_4$ | 0.119 | 0.010 | 1.661 | 0.219 |
| $\mu_5$ | −0.214 | 0.016 | 1.136 | 0.132 | $\psi_5$ | 0.133 | 0.008 | 1.118 | 0.219 |
| $\mu_6$ | −0.144 | 0.023 | 0.773 | 0.113 | $\psi_6$ | 0.421 | 0.021 | 1.718 | 0.199 |
| $\mu_7$ | −0.044 | 0.031 | 0.227 | 0.073 | $\psi_7$ | 0.209 | 0.028 | 0.361 | 0.069 |
| $\mu_8$ | −0.054 | 0.030 | 0.260 | 0.070 | $\psi_8$ | 0.620 | 0.030 | 0.663 | 0.077 |
| $\mu_9$ | −0.085 | 0.026 | 0.499 | 0.099 | $\psi_9$ | 0.282 | 0.024 | 0.823 | 0.253 |
| $\mu_{10}$ | −0.117 | 0.024 | 0.698 | 0.115 | $\psi_{10}$ | 0.486 | 0.026 | 1.941 | 0.293 |
| $\mu_{11}$ | −0.070 | 0.015 | 0.408 | 0.146 | $\psi_{11}$ | 0.079 | 0.007 | 1.721 | 0.494 |
| $\mu_{12}$ | −0.077 | 0.029 | 0.424 | 0.099 | $\psi_{12}$ | 0.507 | 0.026 | 0.946 | 0.127 |
| $\mu_{13}$ | 0.003 | 0.030 | −0.022 | 0.083 | $\psi_{13}$ | 0.139 | 0.012 | 0.334 | 0.054 |
| $\mu_{14}$ | 0.045 | 0.030 | −0.220 | 0.069 | $\psi_{14}$ | 0.294 | 0.017 | 0.328 | 0.046 |
| $\mu_{15}$ | 0.703 | 0.007 | 0.883 | 0.024 | $\psi_{15}$ | 0.034 | 0.002 | 0.095 | 0.010 |
| $\mu_{16}$ | 1.566 | 0.027 | 2.420 | 0.106 | $\psi_{16}$ | 0.121 | 0.011 | 0.581 | 0.119 |
| $\mu_{17}$ | 0.880 | 0.015 | 1.390 | 0.060 | $\psi_{17}$ | 0.099 | 0.006 | 0.430 | 0.049 |
| $\lambda_1$ | 1.182 | 0.046 | 0.852 | 0.078 | $\phi_{11}$ | 0.803 | 0.049 | 0.646 | 0.108 |
| $\lambda_2$ | 1.168 | 0.068 | −2.311 | 0.288 | $\phi_{12}$ | −0.033 | 0.023 | −0.009 | 0.089 |
| $\lambda_3$ | 0.809 | 0.061 | −2.866 | 0.298 | $\phi_{13}$ | −0.251 | 0.022 | −0.360 | 0.080 |
| $\lambda_4$ | 0.917 | 0.080 | −1.664 | 0.242 | $\phi_{14}$ | −0.017 | 0.011 | −0.049 | 0.041 |
| $\lambda_5$ | 0.682 | 0.036 | 0.631 | 0.108 | $\phi_{22}$ | 0.504 | 0.037 | 1.169 | 0.300 |
| $\lambda_6$ | 0.458 | 0.035 | 0.847 | 0.235 | $\phi_{23}$ | 0.247 | 0.020 | 0.108 | 0.099 |
| $\lambda_7$ | 0.628 | 0.028 | 0.164 | 0.183 | $\phi_{24}$ | 0.066 | 0.011 | 0.017 | 0.055 |
| $\lambda_8$ | 1.495 | 0.056 | 0.999 | 0.124 | $\phi_{33}$ | 0.378 | 0.031 | 0.777 | 0.153 |
| $\lambda_9$ | 1.404 | 0.058 | 0.804 | 0.100 | $\phi_{34}$ | 0.044 | 0.008 | 0.083 | 0.048 |
| $\lambda_{10}$ | 0.361 | 0.021 | 0.299 | 0.052 | $\phi_{44}$ | 0.130 | 0.009 | 0.294 | 0.050 |
| $\lambda_{11}$ | 2.126 | 0.070 | 2.373 | 0.215 | $\gamma_1$ | 0.421 | 0.037 | 0.548 | 0.145 |
| $\gamma_2$ | 0.270 | 0.051 | 0.305 | 0.100 | $\gamma_3$ | −0.199 | 0.063 | 0.012 | 0.122 |
| $\gamma_4$ | −0.025 | 0.073 | −0.146 | 0.160 | $\gamma_5$ | 0.111 | 0.027 | 0.169 | 0.059 |
| $\pi$ | 0.845 | 0.012 | 0.155 | 0.012 | b | 0.268 | 0.069 | −0.130 | 0.160 |
| $\psi_{\delta 1}$ | 0.346 | 0.026 | 0.565 | 0.102 | $\psi_{\delta 2}$ | 0.178 | 0.019 | 0.297 | 0.055 |

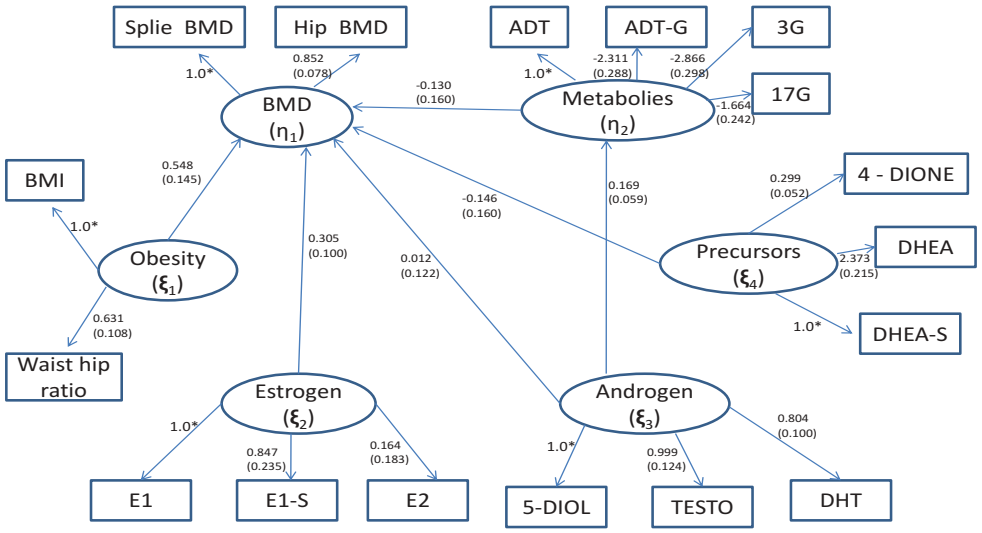
*Note.* Par = parameter; Est = estimate.

We obtain the following observations:

1. The intercepts for both spine BMD and hip BMD are larger in Component 2 than those in Component 1, implying that subjects in Component 2 generally have higher scores (or better conditions) of BMD than those in Component 1.
2. The effect of obesity ($\xi_1$) on BMD ($\eta_1$) is 0.421 (0.037) in Component 1 and 0.548 (0.145) in Component 2. Regardless of components, keeping appropriate weight for old men is beneficial to the prevention of osteoporosis. Likewise, the effect of estrogen ($\xi_2$) on BMD is 0.270 (0.051) in Component 1 and 0.305 (0.100) in Component 2. A




(a) The path diagram and estimation result in component 1.




(b) The path diagram and estimation result in component 2.

FIGURE 6 The proposed model and estimation result in the real example.

high level of estrogen is associated with improved BMD condition. These results are in line with the previous findings (e.g., Lau et al., 2006).

3. The effects of androgen ($\xi_3$) and metabolites ($\eta_2$) on BMD have different patterns; they are −0.199 (0.063) and 0.268 (0.069) in Component 1 and 0.012 (0.122) and −0.130 (0.160) in Component 2, respectively. For those who have a worse condition of BMD, an increase in androgen level or decrease in metabolites accelerates the loss of BMD and raises the risk of osteoporosis. In addition, androgen exhibits significant positive effect on metabolites ($\eta_2$) in both components. This effect is 0.111 (0.027) in Component 1 and 0.169 (0.059) in Component 2. A high level of androgen is associated with a high level of metabolites.
4. Precursors ($\xi_4$) do not significantly affect BMD in both components; the effect is −0.025 (0.073) in Component 1 and −0.146 (0.160) in Component 2. This result agrees with the previous finding discovered by Labrie et al. (2000).

## DISCUSSION

In this article, we developed a full Bayesian approach coupled with the RJMCMC algorithm to analyze mixture SEMs with an unknown number of components. The developed procedure allows the number of components to vary from one MCMC iteration to the next, thereby precluding the necessity to explicitly perform model selection. The results obtained from the simulation study demonstrated that the proposed methodology can accurately select the number of mixture components and simultaneously perform model estimation. The application to the study on the risk factors of osteoporotic fractures provided new insights into osteoporosis prevention and control.

Although available Bayesian model selection methods, such as DIC and Bayes factor, are applicable to the proposed context, they compare models in a pairwise basis. As shown in our simulation study, such pairwise comparison is much more computationally demanding than the proposed procedure, and it will become extremely tedious and time-consuming when the model space increases. By contrast, the proposed RJMCMC procedure is highly computationally efficient because it facilitates an automatic exploration in a large model space and simultaneously compares a number of candidate models in an efficient way. Owing to this appealing feature, the proposed method is recommended, especially in complex situations where the model space is expected to be large.

This study can be extended in several directions: First, given that missing data are frequently encountered in substantive researches, an extension to the case in the presence of ignorable or nonignorable missing data is necessary. Second, hidden Markov models (HMMs) with latent variables are useful multivariate tools to analyze dynamic heterogeneous data. Determining the number of hidden states is likewise a critical issue and has received much attention in the literature (Altman, 2007; Ip, Zhang, Rejeski, Harris, & Kritchevsky, 2013; Maruotti, 2011; Robert et al., 2000; Scott, James, & Suger, 2005; Song, Xia, & Zhu, 2017). However, applying the RJMCMC procedure to simultaneously perform model selection and parameter estimation in the context of HMMs with latent variables requires further investigation. Finally, the development of Bayesian regularized procedures for analyzing mixture models is of future interest.

## ACKNOWLEDGMENTS

The authors thank the editor and the two reviewers for their valuable comments and suggestions, which improved the article substantially.

## FUNDING

This research was supported by GRF 14601115 and 14305014 from the Research Grant Council of the HKSAR, Direct Grants from the Chinese University of Hong Kong, NSFC 11471277 from the National Science Foundation of China, and a grant from Yunnan Province.

## REFERENCES

Altman, R. M. (2007). Mixed hidden Markov models. *Journal of the American Statistical Association*, *102*, 201–210. doi:10.1198/016214506000001086

Arminger, G., & Muthen, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis–Hastings algorithm. *Psychometrika*, *63*, 271–300. doi:10.1007/BF02294856

Arminger, G., Stein, P., & Witterberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika*, *64*, 475–494. doi:10.1007/BF02294568

Bentler, P. M., & Wu, E. (2004). *EQS6: Structural equations program manual*. Encino, CA: Multivariate Software.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley

Boys, R. J., & Henderson, D. A. (2001). A comparison of reversible jump MCMC algorithms for DNA sequence segmentation using hidden Markov models. *Computation Science and Statistics*, *33*, 3549.

Cai, J. H., & Song, X. Y. (2010). Bayesian analysis of mixtures in structural equation models with nonignorable missing data. *British Journal of Mathematical and Statistical Psychology*, *63*, 491–508. doi:10.1348/000711009X475187

Cai, J. H., Song, X. Y., & Hser, Y. I. (2010). A Bayesian analysis of mixture structural equation models with nonignorable missing responses and covariates. *Statistics in Medicine*, *29*, 1861–1874. doi:10.1002/sim.3915

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*, 651–674. doi:10.1214/06-BA122

Chen, J. H., & Khalili, A. (2008). Order selection in finite mixture models with a nonsmooth penalty. *Journal of the American Statistical Association*, *103*, 1674–1683. doi:10.1198/016214508000001075

Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B*, *56*, 363–375.

Dolan, C. V., & van der Mass, H. L. (1998). Fitting multivariate normal mixture subject to structure equation modeling. *Psychometrika*, *63*, 227–253. doi:10.1007/BF02294853

Feng, X. N., Wu, H. T., & Song, X. Y. (2017a). Bayesian adaptive lasso for ordinal regression with latent variables. *Sociological Methods and Research*. doi:10.1177/0049124115610349

Feng, X. N., Wu, H. T., & Song, X. Y. (2017b). Bayesian regularized multivariate generalized latent variable models. *Structural Equation Modeling*, *24*, 341–358. doi:10.1080/10705511.2016.1257353

Fox, E. B., Sudderth, E. B., Jordan, M. I., & Willsky, A. S. (2008). An HDP-HMM for systems with state persistence. In *Proceedings of the 25th International Conference on Machine Learning*, 312–319. ACM Press. doi:10.1145/1390156.1390196

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, *96*, 194–209. doi:10.1198/016214501750333063

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732. doi:10.1093/biomet/82.4.711

Guo, R., Zhu, H., Chow, S. M., & Ibrahim, J. G. (2012). Bayesian lasso for semiparametric structural equation models. *Biometrics*, *68*, 567–577. doi:10.1111/j.1541-0420.2012.01751.x

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chain and their application. *Biometrika*, *57*, 97–109. doi:10.1093/biomet/57.1.97

Huang, T., Peng, H., & Zhang, K. (2017). Model selection for Gaussian mixture models. *Statistica Sinica*, *27*, 147–189.

Ip, E. H., Zhang, Q., Rejeski, W. J., Harris, T. B., & Kritchevsky, S. (2013). Partially ordered mixed hidden Markov model for the disablement process of older adults. *Journal of the American Statistical Association*, *108*, 370–384. doi:10.1080/01621459.2013.770307

Ishwaran, H., & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, *96*, 161–173. doi:10.1198/016214501750332758

Ishwaran, H., & Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, *87*, 371–390. doi:10.1093/biomet/87.2.371

Jasra, A., Holmes, C. C., & Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modelling. *Statistical Science*, *20*, 50–67. doi:10.1214/088342305000000016

Jedidi, K., Jagpal, H., & Desarbo, W. (1997). STEMM: A general finite mixture structural equation model. *Journal of Classification*, *14*, 23–50. doi:10.1007/s003579900002

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. London, UK: Scientific Software International.

Labrie, F., Luu-The, V., Lin, S. X., Simard, J., Labrie, C., El-Alfy, M., … Belanger, A. (2000). Intracrinology: Role of the family of 17 beta-hydroxysteroid dehydrogenases in human physiology and disease. *Journal of Molecular Endocrinology*, *25*, 1–16. doi:10.1677/jme.0.0250001

Lau, E. M., Leung, P. C., Kwok, T., Woo, J., Lynn, H., Orwoll, E., … Cauley, J. (2006). The determinants of bone mineral density in Chinese men—Results from Mr. Os (Hong Kong), the first cohort study on osteoporosis in Asian men. *Osteoporosis International*, *17*, 297–303. doi:10.1007/s00198-005-2019-9

Lee, S. Y. (2007). *Structural equation modelling: A Bayesian approach*. Chichester, UK: Wiley.

Lee, S. Y., & Song, X. Y. (2003). Bayesian model selection for mixture of structure equation models with an unknown number of components. *British Journal of Mathematical and Statistical Psychology*, *56*, 145–165. doi:10.1348/000711003321645403

Li, Y. X., Yutaka, K., Pan, J. H., & Song, X. Y. (2012). A criterion-based model comparison statistic for structural equation models with heterogeneous data. *Journal of Multivariate Analysis*, *112*, 92–107. doi:10.1016/j.jmva.2012.05.010

Lindsay, B. G., & Basak, P. (1993). Multivariate normal mixtures: A fast consistent method of moments. *Journal of the American Statistical Association*, *88*, 468–476. doi:10.1080/01621459.1993.10476297

Maruotti, A. (2011). Mixed hidden Markov models for longitudinal data: An overview. *International Statistical Review*, *79*, 427–454. doi:10.1111/j.1751-5823.2011.00160.x

Matthew, S. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B*, *62*, 795–809.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087–1092. doi:10.1063/1.1699114

Muthén, L. K., & Muthén, B. O. (1998–2007). *M*plus *user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Papastamoulis, P., & Iliopoulos, G. (2009). Reversible jump MCMC in mixtures of normal distributions with the same component means. *Computational Statistics and Data Analysis*, *53*, 900–911. doi:10.1016/j.csda.2008.10.022

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, *9*, 523–539. doi:10.1093/biostatistics/kxm049

Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society, Series B*, *59*, 731–792. doi:10.1111/1467-9868.00095

Robert, C. P., Rydén, T., & Titterington, D. M. (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Series B*, *62*, 57–75. doi:10.1111/1467-9868.00219

Roeder, K., & Wasserman, L. (1997). Practice Bayesian density estimation using mixture of normals. *Journal of the American Statistical Association*, *92*, 894–902. doi:10.1080/01621459.1997.10474044

Scott, S. L., James, G. L., & Suger, C. A. (2005). Hidden Markov models for longitudinal comparisons. *Journal of the American Statistical Association*, *100*, 359–369. doi:10.1198/016214504000001592

Shi, J. Q., & Lee, S. Y. (1998). Bayesian sampling-based approach for factor analysis models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, *51*, 233–252. doi:10.1111/j.2044-8317.1998.tb00679.x

Song, X. Y., & Lee, S. Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences*. London, UK: Wiley.

Song, X. Y., Xia, Y. M., Pan, J. H., & Lee, S. Y. (2011). Model comparison of Bayesian semiparametric and parametric structural equation models. *Structural Equation Modeling*, *18*, 55–72. doi:10.1080/10705511.2011.532720

Song, X. Y., Xia, Y. M., & Zhu, H. T. (2017). Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics*, *73*, 313–323. doi:10.1111/biom.12536

Spezia, L. (2009). Reversible jump and the label switching problem in hidden Markov models. *Journal of Statistical Planning and Inference*, *139*, 2305–2315. doi:10.1016/j.jspi.2008.10.016

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003). *WinBUGS user manual. Version 1.4*. Cambridge, UK: MRC Biostatistics Unit.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association*, *101*, 1566–1581. doi:10.1198/016214506000000302

Titterington, D. M., Smith, A., & Markov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York, NY: Wiley.

Yang, M. G., & Dunson, D. B. (2010). Bayesian semiparametric structural equation models with latent variables. *Psychometrika*, *75*, 675–693. doi:10.1007/s11336-010-9174-4

Yung, Y. F. (1997). Finite mixture in confirmatory factor-analysis models. *Psychometrika*, *62*, 297–330. doi:10.1007/BF02294554

Zhu, H. T., & Lee, S. Y. (2001). A Bayesian analysis of finite mixture in the LISREL model. *Psychometrika*, *66*, 133–152. doi:10.1007/BF02295737