6 Structural Equation Models with Hierarchical and Multisample Data

6.1 Introduction

Structural equation models and Bayesian methods described in previous chapters assume that the available data are obtained from a random sample of a single population. However, in many substantive researches, the related data may exhibit at least two possible kinds of heterogeneity. The first kind is mixture data, which involve independent observations that come from one of the K populations with different distributions, and no information is available on which of the K populations an individual observation belongs to. Although K may be known or unknown, it is usually quite small. Mixture models that will be discussed in Chapter 7 are used to analyze this kind of heterogeneous data. The second kind of heterogeneous data are drawn from a number of different groups (clusters) with a known hierarchical structure. Examples may well be drawing of random samples of patients from within random samples of clinics or hospitals; individuals from within random samples of families; or students from within random samples of schools. In contrast to the mixture data, these hierarchically structured data usually involve a large number of groups, and the group membership of each observation can be exactly specified. However, as individuals within a group share certain common influential factors, the random observations are correlated. Hence, the assumption of independence among observed observations is violated. Clearly, ignoring the correlated structure of the data and analyzing them as observations from a single random sample give erroneous results. Moreover, it is also desirable to establish a meaningful model for the between-group levels, and study the effects of the between-group latent variables on the within-group latent variables.

Multisample data also come from a number of distinct groups (populations). Usually, the number of groups is relatively small, while the number of observations within each group is large. For multisample data, we assume that the observations within each group are independent rather than correlated. Hence, multisample data do not have a hierarchical structure as two-level data. In addition, the number of groups is known, and the group membership of each observation can be specified exactly. In this sense, they can be regarded as a simple mixture data with given correct group label for every observation. As a result, compared to two-level data and/or mixture data, multisample data are easier to cope with.

The objectives of this chapter are to introduce two-level SEMs and multisample SEMs, as well as their associate Bayesian methodologies for analyzing two-level and multisample data.

6.2 Two-level Structural Equation Models

The development of two-level SEMs for taking into consideration the correlated structure of the hierarchical data has received much attention in SEM, see McDonald and Goldstein (1989), Rabe-Hesketh, Skrondal and Pickles (2004), Skrondal and Rabe-Hesketh (2004), and Lee and Song (2005), among others. Using a Bayesian approach, Song and Lee (2004) developed the MCMC methods for analyzing two-level nonlinear models with continuous and ordered categorical data, and Lee, Song and Tang (2007) considered two-level nonlinear SEMs with cross-level effects. For reasons stated in previous chapters, we will describe Bayesian methods for analyzing two-level SEMs in this chapter. To provide a comprehensive framework for analyzing two-level models, nonlinear structural equations are incorporated in the SEMs that are associated with within-group and between-group models. Moreover, the model can accommodate mixed types of continuous and ordered categorical data. In addition to Bayesian estimation, we will present a path sample procedure to compute the Bayes factor for model comparison. The generality of the model is important for providing a comprehensive framework for model comparison of different kinds of SEMs. Again, the idea of data augmentation is utilized. Here, the observed data are augmented with various latent variables at both levels, and the latent continuous random vectors that underlie the ordered categorical variables. An algorithm that is based on the Gibbs sampler (Geman and Geman, 1984) and the Metropolis-Hastings (MH) (Metropolis *et al.*, 1953; Hastings, 1970) algorithm is used for estimation. Observations generated with this algorithm will be used for the path sampling procedure in computing Bayes factor. Although we emphasize on a two-level SEM, the methodology can be extended to higher level SEMs. Finally, an application of WinBUGS to two-level nonlinear SEMs will be discussed.

6.2.1 A Two-level Nonlinear SEM with Mixed Type Variables

Consider a collection of p-variate random vectors $\mathbf{u}_{gi},\ i=1,\cdots,N_g$, nested within groups $g=1,\cdots,G$. The sample sizes N_g may differ from group to group so that the data set is unbalanced. At the first level, we assume that, conditional on the group mean $\mathbf{v}_g$, random observations in each group satisfy the following measurement equation:

$$\mathbf{u}_{gi} = \mathbf{v}_g + \mathbf{\Lambda}_{1g}\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}, \quad g=1,\cdots,G, \ \ i=1,\cdots,N_g, \tag{6.1}$$

where $\mathbf{\Lambda}_{1g}$ is a $p \times q_1$ matrix of factor loadings, $\boldsymbol{\omega}_{1gi}$ is a $q_1 \times 1$ random vector of latent factors, and $\boldsymbol{\epsilon}_{1gi}$ is a $p \times 1$ random vector of error measurements which is independent of $\boldsymbol{\omega}_{1gi}$ and is distributed as $N[\mathbf{0}, \mathbf{\Psi}_{1g}]$, where $\mathbf{\Psi}_{1g}$ is a diagonal matrix. Note that $\mathbf{u}_{gi}$ and $\mathbf{u}_{gj}$ are not independent due to the existence of $\mathbf{v}_g$. Hence, in the two-level SEM, the usual assumption on the independence of the observations is violated. To account for the structure at the between-group level, we assume that the group mean $\mathbf{v}_g$ satisfies the following factor analysis model:

$$\mathbf{v}_g = \boldsymbol{\mu} + \mathbf{\Lambda}_2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}, \quad g=1,\cdots,G, \tag{6.2}$$

where $\boldsymbol{\mu}$ is the vector of intercepts, $\boldsymbol{\Lambda}_2$ is a $p \times q_2$ matrix of factor loadings, $\boldsymbol{\omega}_{2g}$ is a $q_2 \times 1$ vector of latent variables, and $\boldsymbol{\epsilon}_{2g}$ is a $p \times 1$ random vector of error measurements which is independent of $\boldsymbol{\omega}_{2g}$ and is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_2]$, where $\boldsymbol{\Psi}_2$ is a diagonal matrix. Moreover, the first and second level measurement errors are assumed to be independent. It follows from equations (6.1) and (6.2) that

$$\mathbf{u}_{gi} = \boldsymbol{\mu} + \boldsymbol{\Lambda}_2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g} + \boldsymbol{\Lambda}_{1g}\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}. \tag{6.3}$$

In order to assess the interrelationships among the latent variables, latent vectors $\boldsymbol{\omega}_{1gi}$ and $\boldsymbol{\omega}_{2g}$ are partitioned as $\boldsymbol{\omega}_{1gi} = (\boldsymbol{\eta}_{1gi}^T, \boldsymbol{\xi}_{1gi}^T)^T$ and $\boldsymbol{\omega}_{2g} = (\boldsymbol{\eta}_{2g}^T, \boldsymbol{\xi}_{2g}^T)^T$, respectively; where $\boldsymbol{\eta}_{1gi}$ $(q_{11} \times 1)$, $\boldsymbol{\xi}_{1gi}$ $(q_{12} \times 1)$, $\boldsymbol{\eta}_{2g}$ $(q_{21} \times 1)$, and $\boldsymbol{\xi}_{2g}$ $(q_{22} \times 1)$ are latent vectors, with $q_{j1} + q_{j2} = q_j$, for $j = 1,\ 2$. The distributions of $\boldsymbol{\xi}_{1gi}$ and $\boldsymbol{\xi}_{2g}$ are $N[\mathbf{0}, \boldsymbol{\Phi}_{1g}]$ and $N[\mathbf{0}, \boldsymbol{\Phi}_2]$, respectively. The following nonlinear structural equations are incorporated in the between-group and within-group models of the two-level model:

$$\boldsymbol{\eta}_{1gi} = \boldsymbol{\Pi}_{1g}\boldsymbol{\eta}_{1gi} + \boldsymbol{\Gamma}_{1g}\mathbf{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}, \quad \text{and} \tag{6.4}$$

$$\boldsymbol{\eta}_{2g} = \boldsymbol{\Pi}_2\boldsymbol{\eta}_{2g} + \boldsymbol{\Gamma}_2\mathbf{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g}, \tag{6.5}$$

where $\mathbf{F}_1(\boldsymbol{\xi}_{1gi}) = (f_{11}(\boldsymbol{\xi}_{1gi}), \cdots, f_{1a}(\boldsymbol{\xi}_{1gi}))^T$ and $\mathbf{F}_2(\boldsymbol{\xi}_{2g}) = (f_{21}(\boldsymbol{\xi}_{2g}), \cdots, f_{2b}(\boldsymbol{\xi}_{2g}))^T$ are vector-valued functions with nonzero differentiable functions f_{1k} and f_{2k}, and usually $a \geq q_{12}$ and $b \geq q_{22}$, $\boldsymbol{\Pi}_{1g}$ $(q_{11} \times q_{11})$, $\boldsymbol{\Pi}_2$ $(q_{21} \times q_{21})$, $\boldsymbol{\Gamma}_{1g}$ $(q_{11} \times a)$, and $\boldsymbol{\Gamma}_2$ $(q_{21} \times b)$ are unknown parameter matrices, $\boldsymbol{\delta}_{1gi}$ is a vector of error measurements which is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_{1g\delta}]$, $\boldsymbol{\delta}_{2g}$ is a vector of error measurements which is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_{2\delta}]$, and $\boldsymbol{\Psi}_{1g\delta}$ and $\boldsymbol{\Psi}_{2\delta}$ are diagonal matrices. Due to the nonlinearity induced by $\mathbf{F}_1$ and $\mathbf{F}_2$, the underlying distribution of $\mathbf{u}_{gi}$ is not normal. In the within-group structural equation, we assume as usual that $\boldsymbol{\xi}_{1gi}$ and $\boldsymbol{\delta}_{1gi}$ are independent. Similarly, in the between-group structural equation, we assume that $\boldsymbol{\xi}_{2g}$ and $\boldsymbol{\delta}_{2g}$ are independent. Moreover, we assume that the within-group latent vectors $\boldsymbol{\eta}_{1gi}$ and $\boldsymbol{\xi}_{1gi}$ are independent of the between-group latent vectors $\boldsymbol{\eta}_{2g}$ and $\boldsymbol{\xi}_{2g}$. Hence, it follows from (6.4) that $\boldsymbol{\eta}_{1gi}$ is independent of $\boldsymbol{\eta}_{2g}$ and $\boldsymbol{\xi}_{2g}$. That is, this two-level SEM does not accommodate the effects of the latent vectors in the between-group level on the latent vectors in the within-group level (see Lee, 2007, Chapter 9, Section 9.6). However, in the within-group model or in the between-group model, nonlinear effects of explanatory latent variables on outcome latent variables can be assessed through (6.4) and (6.5); and the hierarchical structure of the data has been taken into account. Furthermore, we assume that $\mathbf{I}_1 - \mathbf{\Pi}_{1g}$ and $\mathbf{I}_2 - \mathbf{\Pi}_2$ are nonsingular, and their determinants are independent of the elements in $\mathbf{\Pi}_{1g}$ and $\mathbf{\Pi}_2$, respectively. The two-level SEM is not identified without imposing identification restrictions. The common method of fixing appropriate elements in $\mathbf{\Lambda}_{1g}$, $\mathbf{\Pi}_{1g}$, $\mathbf{\Gamma}_{1g}$, $\mathbf{\Lambda}_2$, $\mathbf{\Pi}_2$, and $\mathbf{\Gamma}_2$ at preassigned known values can be used to achieve an identified model.

To accommodate mixed ordered categorical and continuous variables, without loss of generality, we suppose that $\mathbf{u}_{gi} = (\mathbf{x}_{gi}^T, \mathbf{y}_{gi}^T)^T$, where $\mathbf{x}_{gi} = (x_{gi1}, \cdots, x_{gir})^T$ is a $r \times 1$ observable continuous random vector, and $\mathbf{y}_{gi} = (y_{gi1}, \cdots, y_{gis})^T$ is an $s \times 1$ unobservable continuous random vector. Similar to the previous chapters, a threshold specification is used to model the observable ordered categorical vector $\mathbf{z} = (z_1, \cdots, z_s)^T$ with its underlying continuous vector $\mathbf{y} = (y_1, \cdots, y_s)^T$ as described in (5.3), through integer values in $\{0, 1, \cdots, b_k\}$. Dichotomous variables are treated as an ordered categorical variable with a single threshold that is fixed at zero. The thresholds, mean and variance of an ordered categorical variable can be identified through the method given in Section 5.2.1.

The above model subsumes a number of important models in the recent developments of SEMs. For instance, the models discussed in Shi and Lee (1998, 2000), and Lee and Zhu (2000). Despite its generality, the introduced two-level SEM is defined by measurement and structural equations that describe the relationships among the observed and latent variables at both levels by conceptually simple regression models. Consider the following three major components of the model and the data structure: (1) a two-level model for hierarchically structured data, (2) discrete natures of the data, and (3) a nonlinear structural equation in the within-group model. The first two components are important for achieving correct statistical results. The last component is essential to the analysis of more complicated situations because nonlinear terms of latent variables have been found to be useful in establishing a better model. The between-group model is also defined with a nonlinear structural equation for generality. The development is useful for providing a general framework for analyzing a large number of its submodels. This is particularly true from a model comparison perspective. For example, even if a linear model is better than a nonlinear model in fitting a data set, such a conclusion cannot be reached without the model comparison under the more general nonlinear model framework. For practical situation where G is not large, it may not be worthwhile or practical to consider a complicated between-group model. Moreover, most two-level SEMs in the literature assume that the within-group parameters are invariant over groups.

6.2.2 Bayesian Inference

Motivated by its various advantages, we use the Bayesian approach for analyzing the current two-level nonlinear SEM with mixed continuous, dichotomous, and/or ordered categorical data. Our basic strategy is to augment the observed data with the latent data that consist of the latent variables and/or latent measurements, then MCMC tools are applied to simulate observations in the posterior analysis.

Let $\boldsymbol{\theta}$ be the parameter vector that contains all the unknown structural parameters in $\boldsymbol{\Lambda}_{1g}$, $\boldsymbol{\Psi}_{1g}$, $\boldsymbol{\Pi}_{1g}$, $\boldsymbol{\Gamma}_{1g}$, $\boldsymbol{\Phi}_{1g}$, $\boldsymbol{\Psi}_{1g\delta}$, $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Psi}_2$, $\boldsymbol{\Pi}_2$, $\boldsymbol{\Gamma}_2$, $\boldsymbol{\Phi}_2$, and $\boldsymbol{\Psi}_{2\delta}$, and $\boldsymbol{\alpha}$ be the parameter vector that contains all the unknown thresholds. The total number of unknown parameters in $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ is usually large. In the following analysis, we assume that the two-level nonlinear model defined by $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ is identified. Let $\mathbf{X}_g = (\mathbf{x}_{g1}, \cdots, \mathbf{x}_{gNg})$ and $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_G)$ be the observed continuous data, $\mathbf{Z}_g = (\mathbf{z}_{g1}, \cdots, \mathbf{z}_{gNg})$ and $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_G)$ be the observed ordered categorical data. Let $\mathbf{Y}_g = (\mathbf{y}_{g1}, \cdots, \mathbf{y}_{gNg})$ and $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_G)$ be the latent continuous measurements associated with $\mathbf{Z}_g$ and $\mathbf{Z}$, respectively. The observed data will be augmented with $\mathbf{Y}$ in the posterior analysis. Once $\mathbf{Y}$ is given, all the data are continuous and the problem will be easier to cope with. Let $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_G)$ be the matrix of between-group latent variables. If $\mathbf{V}$ is observed, the model is reduced to the single level multi-sample model. Moreover, let $\boldsymbol{\Omega}_{1g} = (\boldsymbol{\omega}_{1g1}, \cdots, \boldsymbol{\omega}_{1gNg}), \boldsymbol{\Omega}_1 = (\boldsymbol{\Omega}_{11}, \cdots, \boldsymbol{\Omega}_{1G})$ and $\boldsymbol{\Omega}_2 = (\boldsymbol{\omega}_{21}, \cdots, \boldsymbol{\omega}_{2G})$ be the matrices of latent variables at the within-group and between-group levels. If these matrices are observed, the complicated nonlinear structural equations (6.4) and (6.5) reduce to the regular simultaneous regression model. Difficulties due to the nonlinear relationships among the latent variables are greatly alleviated. Hence, problems associated with the complicated components of the model, such as the correlated structure of the observations induced by the two-level data, discrete nature of the ordered categorical variables, and the nonlinearity of the latent variables at both levels, can be handled by data augmentation. In the posterior analysis, the observed data $(\mathbf{X}, \mathbf{Z})$ will be augmented with $(\mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$, and we will consider the joint posterior distribution $[\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2 | \mathbf{X}, \mathbf{Z}]$. The Gibbs sampler (Geman and Geman, 1984) will be used to generate a sequence of observations from this joint posterior distribution. The Bayesian solution is then obtained using standard inferences based on the generated sample of observations. In applying the Gibbs sampler, we iteratively sample from the following conditional distributions: $[\mathbf{V} | \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$, $[\boldsymbol{\Omega}_1 | \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$, $[\boldsymbol{\Omega}_2 | \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \mathbf{X}, \mathbf{Z}]$, $[\boldsymbol{\alpha}, \mathbf{Y} | \boldsymbol{\theta}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$, and $[\boldsymbol{\theta} | \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$.

For the defined two-level model, the conditional distribution $[\boldsymbol{\theta} | \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$

is further decomposed into components involving various structural parameters in the between-group and within-group models. These components are different under various special cases of the model. Some typical examples are:

(A) Models with different within-group parameters across groups: In this case, the within-group structural parameters $\boldsymbol{\theta}_{1g} = \{\boldsymbol{\Lambda}_{1g}, \boldsymbol{\Psi}_{1g}, \boldsymbol{\Pi}_{1g}, \boldsymbol{\Gamma}_{1g}, \boldsymbol{\Phi}_{1g}, \boldsymbol{\Psi}_{1g\delta}\}$ and threshold parameters $\boldsymbol{\alpha}_g$ associated with the gth group are different from those associated with the hth group, for $g \neq h$. Practically, G and N_g should not be too small for drawing valid statistical conclusions for the between-group model and the gth within-group model.

(B) Models with some invariant within-group parameters: In this case, parameters $\boldsymbol{\theta}_{1g}$ and/or $\boldsymbol{\alpha}_g$ associated with the gth group are equal to those associated with some other groups.

(C) Models with all invariant within-group parameters: Under this situation, $\boldsymbol{\theta}_{11} = \cdots = \boldsymbol{\theta}_{1G}$, and $\boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_G$.

Conditional distributions under various special cases are similar but different. Moreover, prior distributions of the parameters are also involved. On the basis of the reasoning given in previous chapters, conjugate type prior distributions are used. The non-informative distribution is used for the prior distribution of the thresholds. The conditional distributions of the components in $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$ as well as other conditional distributions required by the Gibbs sampler are briefly discussed in Appendix 6.1. As we can see in this appendix, these conditional distributions are generalizations of those that are associated with a single level model; and most of them are standard distributions such as normal, Gamma, and inverted Wishart distributions. Simulating observations from them requires little computing time. The MH algorithm will be used for simulating observations efficiently from the following three complicated conditional distributions, $[\boldsymbol{\Omega}_1|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$, $[\boldsymbol{\Omega}_2|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \mathbf{X}, \mathbf{Z}]$, and $[\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$. Some technical details on the implementation of the MH algorithm are given in Appendix 6.2.

The following Bayes factor (see Kass and Raftery, 1995) is used for comparing competing M_0 and M_1:

$$B_{10} = \frac{p(\mathbf{X}, \mathbf{Z}|M_1)}{p(\mathbf{X}, \mathbf{Z}|M_0)}.$$

In the application of path sampling in computing B_{10}, we again use the data augmentation idea to augment $(\mathbf{X}, \mathbf{Z})$ with $(\mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$ in the analysis. It can be shown by similar reasoning as in Chapter 4 that:

$$\widehat{\log B_{10}} = \frac{1}{2}\sum_{s=0}^{S}(t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_{(s)}), \tag{6.6}$$

where $0 = t_{(0)} < t_{(1)} < \cdots < t_{(S)} < t_{(S+1)} = 1$ are fixed grids in $[0, 1]$ and

$$\bar{U}_{(s)} = \frac{1}{J}\sum_{j=1}^{J} U(\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{V}^{(j)}, \boldsymbol{\Omega}_1^{(j)}, \boldsymbol{\Omega}_2^{(j)}, \mathbf{X}, \mathbf{Z}, t_{(s)}), \tag{6.7}$$

in which $\{(\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{V}^{(j)}, \boldsymbol{\Omega}_1^{(j)}, \boldsymbol{\Omega}_2^{(j)}),\ j = 1, \cdots, J\}$ is a sample of observations simulated from the joint posterior distribution $[\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2|\mathbf{X}, \mathbf{Z}, t_{(s)}]$, and

$$U(\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}, t) = d \log p(\mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\alpha}, t)/dt, \tag{6.8}$$

where $p(\mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\alpha}, t)$ is the complete-data likelihood. Note that this complete-data likelihood is not complicated, thus obtaining the function $U(\cdot)$ through differentiation is not difficult. Moreover, the program implemented in estimation can be used for simulating observations in Equation (6.7), hence little additional programming effort is required. Usually, $S = 10$ grids is sufficient for providing a good approximation of the logarithm B_{10} for not far apart competing models. More grids are required for very different M_1 and M_0, and the issue should be approached on a problem-by-problem basis. In Equation (6.7), a value of $J = 2,000$ is commonly enough for most practical applications.

An important step in applying path sampling for computing logarithm B_{12} is to find a good path t in $[0, 1]$ to link the competing models M_1 and M_2. Because the two-level nonlinear SEM is rather complex, M_1 and M_2 can be quite different and finding a path to link them may require some insight. An illustrative example is discussed as follows.

Example: The competing models M_1 and M_2 have the following within-group measurement and structural equations:

$$\mathbf{u}_{gi} = \mathbf{v}_g + \mathbf{\Lambda}_1\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}, \tag{6.9}$$

$$\boldsymbol{\eta}_{1gi} = \mathbf{\Pi}_1\boldsymbol{\eta}_{1gi} + \mathbf{\Gamma}_1\mathbf{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}. \tag{6.10}$$

The difference between M_1 and M_2 is on the between-group models. Let

$$M_1: \ \mathbf{v}_g = \boldsymbol{\mu} + \mathbf{\Lambda}_2^1\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}, \tag{6.11}$$

where $\boldsymbol{\omega}_{2g}$ is distributed as $N[\mathbf{0}, \mathbf{\Phi}_2]$. Thus, the between-group model in M_1 is a factor analysis model. In M_2, $\boldsymbol{\omega}_{2g} = (\boldsymbol{\eta}_{2g}^T, \boldsymbol{\xi}_{2g}^T)^T$, and the measurement and structural equations in the between-group model are given as follows:

$$M_2: \quad \mathbf{v}_g = \boldsymbol{\mu} + \mathbf{\Lambda}_2^2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}, \tag{6.12}$$

$$\boldsymbol{\eta}_{2g} = \mathbf{\Pi}_2^2\boldsymbol{\eta}_{2g} + \mathbf{\Gamma}_2^2\mathbf{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g}. \tag{6.13}$$

The between-group model in M_2 is a nonlinear SEM with a nonlinear structural equation. Note that M_1 and M_2 are non-nested. As there are two different models for $\boldsymbol{\omega}_{2g}$, it is rather difficult to directly link M_1 and M_2. This difficulty can be solved via an auxiliary model M_a which can be linked with both M_1 and M_2. We first compute log B_{1a} and log B_{2a}, and then obtain $\log B_{12}$ via the following equation:

$$\log B_{12} = \log \frac{p(\mathbf{X}, \mathbf{Z}|M_1)/p(\mathbf{X}, \mathbf{Z}|M_a)}{p(\mathbf{X}, \mathbf{Z}|M_2)/p(\mathbf{X}, \mathbf{Z}|M_a)} = \log B_{1a} - \log B_{2a}. \tag{6.14}$$

For the above problem, one auxiliary model is M_a, in which the measurement and structural equations of the within-group model are given by (6.9) and (6.10), and the between-group model is defined by $\mathbf{v}_g = \boldsymbol{\mu} + \boldsymbol{\epsilon}_{2g}$. The link model M_{t1a} is defined by

M_{t1a} : $\mathbf{u}_{gi} = \boldsymbol{\mu} + t\boldsymbol{\Lambda}_2^1\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g} + \boldsymbol{\Lambda}_1\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}$, with the within-group structural equation given by (6.10), where $\boldsymbol{\omega}_{2g}$ is distributed as $N[\mathbf{0}, \boldsymbol{\Phi}_2]$ and without a between-group structural equation. Clearly, $t = 1$ and 0 corresponds to M_1 and M_a, respectively. Hence, $\log B_{1a}$ can be computed under this setting via the path sampling procedure. The link model M_{t2a} is defined by $M_{t2a} : \mathbf{u}_{gi} = \boldsymbol{\mu} + t\boldsymbol{\Lambda}_2^2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g} + \boldsymbol{\Lambda}_1\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}$, with the within-group and between-group structural equations given by (6.10) and (6.13), respectively. Clearly, $t = 1$ and 0 corresponds to M_2 and M_a. Hence, $\log B_{2a}$ can be obtained. Finally, $\log B_{12}$ can be obtained from $\log B_{1a}$ and $\log B_{2a}$ via (6.14).

In general, just one auxiliary model may not be adequate to link two very different models M_1 and M_2. However, based on the idea of the above example, the difficulty can be solved by using more than one appropriate auxiliary models $M_a,\ M_b, \cdots$ between M_1 and M_2. For example, suppose we use M_a and M_b to link M_1 and M_2, with M_a closer to M_1 and M_b closer to M_2. Then

$$\frac{p(\mathbf{X},\mathbf{Z}|M_1)}{p(\mathbf{X},\mathbf{Z}|M_2)} = \frac{p(\mathbf{X},\mathbf{Z}|M_1)/p(\mathbf{X},\mathbf{Z}|M_a)}{p(\mathbf{X},\mathbf{Z}|M_2)/p(\mathbf{X},\mathbf{Z}|M_a)}, \quad \text{and} \quad \frac{p(\mathbf{X},\mathbf{Z}|M_2)}{p(\mathbf{X},\mathbf{Z}|M_a)} = \frac{p(\mathbf{X},\mathbf{Z}|M_2)/p(\mathbf{X},\mathbf{Z}|M_b)}{p(\mathbf{X},\mathbf{Z}|M_a)/p(\mathbf{X},\mathbf{Z}|M_b)};$$

hence, $\log B_{12} = \log B_{1a} + \log B_{ab} - \log B_{2b}$. Each logarithm Bayes factor can be computed via path sampling.

Similar to the goodness-of-fit assessment in the context of single level nonlinear SEMs, it is rather difficult to find a saturated model for the two-level nonlinear SEMs. However, the goodness-of-fit of a proposed model can be assessed by means of the posterior predictive (PP) p-value and the estimated residual plots. The PP p-value (Gelman, Meng and Stern, 1996) can be used as a goodness-of-fit assessment for a hypothesized two-level nonlinear SEM with the mixed types data. A brief description of the PP p-values in the context of the current model is given in Appendix 6.3.

6.2.3 An Application: Filipina CSWs Study

As an illustration of the methodology, we use a small portion of the data set in the study of Morisky *et al.* (1998) on the effects of establishment policies, knowledge, and attitudes on condom use among Filipina commercial sex workers (CSWs), see Morisky *et al.* (1998). As commercial sex work promotes the spread of AIDS and other sexually transmitted diseases; promotion of safer sexual practice among CSWs is important. The study of Morisky *et al.* (1998) concerned the development and preliminary findings from an AIDS preventative intervention for Filipina CSWs. The data set was collected from female CSWs in establishments (bars, night clubs, Karaoke TV, and massage parlous) in cities of Philippines. The whole questionnaire consisted of 134 items on areas of demographics knowledge, attitudes, beliefs, behaviors, self-efficacy for condom use, and social desirability. Latent psychological determinants such as CSWs' risk behaviors, knowledge and attitudes associated with AIDS and condom use are important issues to be assessed. For instance, a basic concern is to explore whether linear relationships among these latent variables are sufficient, or it is better to incorporate nonlinear relationships in the model. The observed variables that are used as indicators for latent quantities are measured in terms of ordered categorical and continuous scales. Moreover, as emphasized by Morisky *et al.* (1998), establishment policies on their CSWs condom use practices exert a strong influence on CSWs. Hence, it is interesting to study the influence of the establishment by incorporating a between-group model for the data. As observations within each establishment are correlated, the usual assumption of independence in the standard single-level SEMs is violated. On the basis of the above considerations, it is desirable to employ a two-level nonlinear SEM in the context of mixed ordered categorical and continuous data.

Nine observed variables, of which the 7th, 8th, and 9th variables are continuous and the remaining are ordered categorical with a five-point scale, are selected. Questions corresponding to these variables are given in Appendix 1.1 (Data set (iv), Questions (1)-(9)). For brevity, we delete those observations with missing entries in the analysis, and the remaining sample size is 755. There are 97 establishments. The numbers of individuals in establishments varied from 1 to 58; this gives an unbalanced data set. The sample means and standard deviations of the continuous variables are $\{2.442, 1.180, 0.465\}$ and $\{5.299, 2.208, 1.590\}$, respectively. The cell frequencies of the ordered categorical variables vary from 12 to 348. To unify scales of variables, the raw continuous data are standardized.

After some preliminary studies and based on the meanings of the questions corresponding to the observed variables (see Appendix 1.1), in the measurement equations corresponding to the between-group and within-group models, we use the first three, the next three, and the last three observed variables as indicators for latent factors that can be roughly interpreted as 'worry about AIDS', 'attitude to the risk of getting AIDS', and 'aggressiveness'. For the between-group model, we use a factor analysis model with the following specifications:

$$\mathbf{\Lambda}_2^T = \begin{bmatrix} 1 & \lambda_{2,21} & \lambda_{2,31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda_{2,52} & \lambda_{2,62} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{2,83} & \lambda_{2,93} \end{bmatrix},$$

$$\mathbf{\Phi}_2 = \begin{bmatrix} \phi_{2,11} & & \text{sym} \\ \phi_{2,21} & \phi_{2,22} & \\ \phi_{2,31} & \phi_{2,32} & \phi_{2,33} \end{bmatrix}$$

and $\mathbf{\Psi}_2 = \text{diag}(0.3, 0.3, 0.3, 0.3, 0.3, 0.3, \psi_{27}, \psi_{28}, \psi_{29})$, where the unique variances corresponding to the ordered categorical variables are fixed at 0.3. Although other structure for $\mathbf{\Lambda}_2$ can be considered, we choose this common form in a confirmatory factor analysis that gives non-overlapping latent factors for clear interpretation. These latent factors are allowed to be correlated. For the within-group model with the latent factors $\{\eta_{1gi}, \xi_{1gi1}, \xi_{1gi2}\}$, we considered invariant within-group parameters such that $\mathbf{\Psi}_{1g} = \mathbf{\Psi}_1 = \mathrm{diag}(\psi_{11}, \cdots, \psi_{19})$, and $\mathbf{\Lambda}_{1g} = \mathbf{\Lambda}_1$, where $\mathbf{\Lambda}_1$ has the same common structure as $\mathbf{\Lambda}_2$ with unknown loadings $\{\lambda_{1,21}, \lambda_{1,31}, \lambda_{1,52}, \lambda_{1,62}, \lambda_{1,83}, \lambda_{1,93}\}$. However, as the within-group model is directly related to the CSWs, we wish to consider a more subtle model with a structural equation that accounts for relationships among the latent factors. To assess the interaction effect of the explanatory latent factors, the following structural equation for the latent variables is taken:

$$\eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{13}\xi_{1gi1}\xi_{1gi2} + \delta_{1gi}. \tag{6.15}$$

To identify the model with respect to ordered categorical variables via the common method, α_{k1} and $\alpha_{k4},\ k = 1, \cdots, 6$ are fixed at $\alpha_{kj} = \Phi^{*^{-1}}(m_k)$, where Φ^* is the distribution function of $N[0, 1]$, and m_k is the observed cumulative marginal proportion of the categories with $z_{gk} < j$. There are a total of 49 parameters in this two-level nonlinear SEM.

In the Bayesian analysis, we need to specify hyperparameter values in the proper conjugate prior distributions of the unknown parameters. In this illustrative example, we use some data-dependent prior inputs, and ad hoc prior inputs that give rather vague but proper prior distributions. We emphasize that these prior inputs are used for the purpose of illustration only, we are not routinely recommending them for other substantive applications. The data-dependent prior inputs are obtained by conducting an auxiliary Bayesian estimation with proper vague conjugate prior distributions which gives estimates $\tilde{\mathbf{\Lambda}}_{01k}$, $\tilde{\mathbf{\Lambda}}^*_{01k}$, $\tilde{\mathbf{\Lambda}}_{02k}$, and $\tilde{\mathbf{\Lambda}}^*_{02k}$ for some hyperparameter values (according to the notation in Appendix 6.1). Then, results are obtained and compared on the basis of the following types of hyperparameter values:

(I): Hyperparameters $\mathbf{\Lambda}_{01k}$, $\mathbf{\Lambda}^*_{01k}$, $\mathbf{\Lambda}_{02k}$, and $\mathbf{\Lambda}^*_{02k}$ are equal to $\tilde{\mathbf{\Lambda}}_{01k}$, $\tilde{\mathbf{\Lambda}}^*_{01k}$, $\tilde{\mathbf{\Lambda}}_{02k}$, and $\tilde{\mathbf{\Lambda}}^*_{02k}$, respectively; $\mathbf{H}_{01k}$, $\mathbf{H}_{01k}^*$, $\mathbf{H}_{02k}$, and $\mathbf{H}_{02k}^*$ are equal to identity matrices of appropriate orders; $\alpha_{01k} = \alpha_{02k} = \alpha_{01\delta k} = \alpha_{02\delta k} = 10$, $\beta_{01k} = \beta_{02k} = \beta_{01\delta k} = \beta_{02\delta k} = 8$, $\rho_{01} = \rho_{02} = 6$, $\mathbf{R}_{01}^{-1} = 5\mathbf{I}_2$, and $\mathbf{R}_{02}^{-1} = 5\mathbf{I}_3$.

(II): Hyperparameter values in $\mathbf{\Lambda}_{01k}$, $\mathbf{\Lambda}_{01k}^*$, $\mathbf{\Lambda}_{02k}$, and $\mathbf{\Lambda}_{02k}^*$ are equal to zeros, $\mathbf{H}_{01k}$, $\mathbf{H}_{01k}^*$, $\mathbf{H}_{02k}$, and $\mathbf{H}_{02k}^*$ are equal to 5 times the identity matrices of appropriate orders. Other hyperparameter values are equal to those given in (I). These prior inputs are not data-dependent.

Bayesian estimates are obtained using the MCMC algorithm that includes the Gibbs sampler and the MH algorithm. The convergence of this algorithm is monitored by plots of generated observations obtained with different starting values. We observe that the algorithm converges in less than 2,000 iterations. Hence, we take a burn-in-phase of 2,000 iterations, and further collect 3,000 observations to produce the Bayesian estimates and their standard error estimates. Results obtained under prior inputs (I) and (II) are reported in Tables 6.1 and 6.2, respectively. We see that the estimates obtained under these different prior inputs are reasonably close. The PP p-values corresponding to these two sets of estimates are equal to 0.592 and 0.600, indicating that the proposed model fit the sample data, and that this statistic is quite robust to the selected prior inputs under the given sample size of 755.

Tables 6.1 and 6.2 here

In order to illustrate the path sampling in computing the Bayes factor for model comparison, we compare this two-level nonlinear model with some non-nested models. Let M_1 be the two-level nonlinear SEM with the above specifications and the nonlinear structural equation given in (6.15), and M_2 and M_3 be non-nested models with the same specifications except that the corresponding nonlinear structural equations are given by

$$M_2: \quad \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{14}\xi_{1gi1}^2 + \delta_{1gi}, \tag{6.16}$$

$\sum_{i=1}\sum_{k=1}^{p} \exp\left\{ \frac{(y_{ik}-\mu_{\cdot}\cdots)^T(\quad}{\psi_{\epsilon k}} \right\}$ $\quad \varepsilon_i^{(g)} = v_i^{(g)} - [\mu^{(g)} + (1-t)\Lambda\omega_i^{(g)} + t\Lambda^{(g)}\omega_i^{(g)}]$,

$\sim N(0, \Psi_\varepsilon)$

$$M_3: \quad \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{15}\xi_{1gi2}^2 + \delta_{1gi}. \tag{6.17}$$

To apply the path sampling to compute the Bayes factor for comparing M_1 and M_2, we link up M_1 and M_2 by M_t with the following structural equation:

$$\eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + (1-t)\gamma_{13}\xi_{1gi1}\xi_{1gi2} + t\gamma_{14}\xi_{1gi1}^2 + \delta_{1gi}. \tag{6.18}$$

Clearly, when $t = 1$, $M_t = M_2$; when $t = 0$, $M_t = M_0$. By differentiating the complete-data log-likelihood $\log p(\mathbf{Y}, \mathbf{V}, \mathbf{\Omega}_1, \mathbf{\Omega}_2, \mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\alpha}, t)$ with respect to t, we obtain

$$U(\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \mathbf{\Omega}_1, \mathbf{\Omega}_2, \mathbf{X}, \mathbf{Z}, t) = \sum_{i=1}^{n}\{\eta_{1gi} - \gamma_{11}\xi_{1gi1} - \gamma_{12}\xi_{1gi2}$$
$$- t\gamma_{13}\xi_{1gi1}\xi_{1gi2} - (1-t)\gamma_{14}\xi_{1gi1}^2\}\psi_{1\delta}^{-1}\{\gamma_{13}\xi_{1gi1}\xi_{1gi2} - \gamma_{14}\xi_{1gi1}^2\}.$$

Consequently, $\log B_{12}$ can be computed using (6.6) and (6.7) with a sample of observations simulated from the appropriate posterior distributions. The above procedure can be similarly used for computing $\log B_{13}$ and $\log B_{23}$.

In this example, we take 20 grids in [0,1] and $J = 1,000$ in computing logarithm Bayes factors. The estimated $\log B_{21}$ and $\log B_{23}$ under prior inputs (I, II) are equal to (0.317, 0.018) and (0.176, 0.131), respectively. Hence, the values of logarithm Bayes factors are reasonably close with the given different prior inputs. According to the criterion given in Kass and Raftery (1995) for comparing non-nested models, M_2 is slightly better than M_1 and M_3. To apply the procedure for comparing nested models, we further compare M_2 with a linear model M_0 and a more comprehensive model M_4. Competing models M_0 and M_4 have the same specifications as M_2, except the corresponding structural equations are given by:

$$M_0: \quad \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \delta_{1gi},$$

$$M_4: \quad \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{13}\xi_{1gi1}\xi_{1gi2} + \gamma_{14}\xi_{1gi1}^2 + \gamma_{15}\xi_{1gi2}^2 + \delta_{1gi}.$$

Note that M_0 is nested in M_2, and M_2 is nested in M_4. Using the path sampling procedure, the estimated $\log B_{40}$ and $\log B_{42}$ under prior inputs (I, II) are (1.181, 1.233) and (1.043, 1.071), respectively. Hence, M_4 is better than M_0 and M_2. The PP p-values corresponding to M_4 under prior inputs (I) and (II) are equal to 0.582 and 0.611, respectively. These two values are close and indicate the expected result that the selected model also fits the data. The Bayesian estimates and their standard error estimates under M_4 and Prior inputs (I) are reported in Table 6.3. Results obtained under prior inputs (II) are similar. We also observe comparatively large variabilities for estimates of parameters in $\mathbf{\Lambda}_2$ corresponding to the between-group model and estimates of $\{\gamma_{13}, \gamma_{14}, \gamma_{15}\}$ corresponding to the nonlinear terms of the latent variables. This phenomenon may be due to the small sample size at the between-group level and the complicated nature of the parameters. Other straightforward interpretations are not presented. Based on the abovementioned methodology, more complicated or other combinations of nonlinear terms can be analyzed similarly.

To summarize, we have introduced a two-level nonlinear SEM with three non-overlapping factors: 'worry about AIDS', 'attitude to the risk of getting AIDS', and 'aggressiveness' in the within-group and between-group covariance structures. The significance of the establishments influence is reflected by relatively large estimates of some between-group parameters.

The software WinBUGS (Spiegelhalter *et al.*, 2003) can produce Bayesian estimates of the parameters in some two-level nonlinear SEMs. To demonstrate this, we apply WinBUGS to analyze the current AIDS data based on M_4 and Type I prior inputs. The WinBUGS code is given in Appendix 6.4.

Table 6.3 here

6.3 Structural Equation Models with Multisample Data

In contrast to multilevel data, multisample data come from a comparatively smaller number of groups (populations); the number of observations within each group is usually large, and observations within each group are assumed independent. One main objective in the analysis of multisample data is to investigate the similarities or differences among the models in the different groups. As a result, the statistical inferences emphasized in analyzing multisample SEMs are different from those in analyzing two-level SEMs. Analysis of multiple samples is a major topic in structural equation modeling. It is useful for investigating the behaviors of different groups of employees, different cultures, and different treatment groups, etc. The main interest is the testing of hypotheses about the different kinds of invariances among the models in different groups. This issue can be formulated as a model comparison problem, and can be effectively addressed by the Bayes factor or DIC in a Bayesian approach. Another advantage of the Bayesian model comparison through the Bayes factor or DIC is that nonnested models (hypotheses) can be compared, hence it is not necessary to follow a hierarchy of hypotheses to assess the invariance for the SEMs in different groups.

Bayesian methods for analyzing mutlisample SEMs are presented in this section. We will emphasize on nonlinear SEMs with ordered categorical variables, although the general ideas can be applied to SEMs with other settings.

6.3.1 Bayesian Analysis of a Nonlinear SEM in Different Groups

Consider G independent groups of individuals that represent different populations. For $g = 1, \cdots, G$, and $i = 1, \cdots, N_g$, let $\mathbf{v}_i^{(g)}$ be the $p \times 1$ random vector of observed variables that correspond to the ith observation (subject) in the gth group. In contrast to two-level SEMs, for $i = 1, \cdots, N_g$ in the gth group, $\mathbf{v}_i^{(g)}$ are assumed to be independent. For each $g = 1, \cdots, G$, $\mathbf{v}_i^{(g)}$ is related to latent variables in a $q \times 1$ random vector $\boldsymbol{\omega}_i^{(g)}$

through the following measurement equation:

$$\mathbf{v}_i^{(g)} = \boldsymbol{\mu}^{(g)} + \boldsymbol{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \tag{6.19}$$

where $\boldsymbol{\mu}^{(g)}$, $\boldsymbol{\Lambda}^{(g)}$, $\boldsymbol{\omega}_i^{(g)}$, and $\boldsymbol{\epsilon}_i^{(g)}$ are similarly defined as before. It is assumed that $\boldsymbol{\omega}_i^{(g)}$ and $\boldsymbol{\epsilon}_i^{(g)}$ are independent, and the distribution of $\boldsymbol{\epsilon}_i^{(g)}$ is $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon^{(g)}]$, where $\boldsymbol{\Psi}_\epsilon^{(g)}$ is diagonal. Let $\boldsymbol{\omega}_i^{(g)} = (\boldsymbol{\eta}_i^{(g)T}, \boldsymbol{\xi}_i^{(g)T})^T$. It is naturally assumed that the dimensions of $\boldsymbol{\xi}_i^{(g)}$ and $\boldsymbol{\eta}_i^{(g)}$ are independent of g; that is, they are the same for each group. To assess the effects of the nonlinear terms of latent variables in $\boldsymbol{\xi}_i^{(g)}(q_2 \times 1)$ on $\boldsymbol{\eta}_i^{(g)}(q_1 \times 1)$, we consider a nonlinear SEM with the following nonlinear structural equation:

$$\boldsymbol{\eta}_i^{(g)} = \boldsymbol{\Pi}^{(g)}\boldsymbol{\eta}_i^{(g)} + \boldsymbol{\Gamma}^{(g)}\mathbf{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\delta}_i^{(g)}, \tag{6.20}$$

where $\boldsymbol{\Pi}^{(g)}$, $\boldsymbol{\Gamma}^{(g)}$, $\mathbf{F}(\boldsymbol{\xi}_i^{(g)})$, and $\boldsymbol{\delta}_i^{(g)}$ are similarly defined as before. It is assumed that $\boldsymbol{\xi}_i^{(g)}$ and $\boldsymbol{\delta}_i^{(g)}$ are independent, the distributions of $\boldsymbol{\xi}_i^{(g)}$ and $\boldsymbol{\delta}_i^{(g)}$ are $N[\mathbf{0}, \boldsymbol{\Phi}^{(g)}]$ and $N[\mathbf{0}, \boldsymbol{\Psi}_\delta^{(g)}]$, respectively, where $\boldsymbol{\Psi}_\delta^{(g)}$ is diagonal; and the vector-valued function $\mathbf{F}(\cdot)$ does not depend on g. However, different groups can have different linear or nonlinear terms of $\boldsymbol{\xi}_i^{(g)}$ by defining appropriate $\mathbf{F}(\cdot)$ and assigning zero values to appropriate elements in $\boldsymbol{\Gamma}^{(g)}$. Let $\boldsymbol{\Lambda}_\omega^{(g)} = (\boldsymbol{\Pi}^{(g)}, \boldsymbol{\Gamma}^{(g)})$ and $\mathbf{G}(\boldsymbol{\omega}_i^{(g)}) = (\boldsymbol{\eta}_i^{(g)^T}, \mathbf{F}(\boldsymbol{\xi}_i^{(g)})^T)^T$, Equation (6.20) can then be rewritten as

$$\boldsymbol{\eta}_i^{(g)} = \boldsymbol{\Lambda}_\omega^{(g)}\mathbf{G}(\boldsymbol{\omega}_i^{(g)}) + \boldsymbol{\delta}_i^{(g)}. \tag{6.21}$$

To handle the ordered categorical outcomes, suppose that $\mathbf{v}_i^{(g)} = (\mathbf{x}_i^{(g)T}, \mathbf{y}_i^{(g)T})^T$, where $\mathbf{x}_i^{(g)}$ is a $r \times 1$ subvector of observable continuous responses; while $\mathbf{y}_i^{(g)}$ is an $s \times 1$ subvector of unobservable continuous responses, the information of which is reflected by an observable ordered categorical vector $\mathbf{z}_i^{(g)}$. In a generic sense, an ordered categorical variable $z_m^{(g)}$ is defined with its underlying latent continuous random variable $y_m^{(g)}$ by:

$$z_m^{(g)} = a \quad \text{if} \quad \alpha_{m,a}^{(g)} \le y_m^{(g)} < \alpha_{m,a+1}^{(g)}, \quad a = 0, \cdots, b_m, \quad m = 1, \cdots, s, \tag{6.22}$$

where $\{-\infty = \alpha_{m,0}^{(g)} < \alpha_{m,1}^{(g)} < \cdots < \alpha_{m,b_m}^{(g)} < \alpha_{m,b_m+1}^{(g)} = \infty\}$ is the set of threshold parameters that define the categories, and $b_m + 1$ is the number of categories for the ordered categorical variable $z_m^{(g)}$. For each ordered categorical variable, the number of thresholds is the same for each group. To tackle the identification problem related to the ordered categorical variables, we consider an ordered categorical variable $z_m^{(g)}$ that is defined by a set of thresholds $\alpha_{m,k}^{(g)}$ and an underlying latent continuous variable $y_m^{(g)}$ with a distribution $N[\mu_m^{(g)}, \sigma_m^{2(g)}]$. The indeterminacy is caused by the fact that $\alpha_{m,k}^{(g)}$, $\mu_m^{(g)}$, and $\sigma_m^{2(g)}$ are not simultaneously estimable. For a given group g, a common method to solve this identification problem with respect to each mth ordered categorical variable that corresponds to the gth group is to fix $\alpha_{m,1}^{(g)}$ and $\alpha_{m,b_m}^{(g)}$ at preassigned values (see Lee *et al.*, 2005). For example, we may fix $\alpha_{m,1}^{(g)} = \Phi^{*-1}(f_{m,1}^{(g)})$, and $\alpha_{m,b_m}^{(g)} = \Phi^{*-1}(f_{m,b_m}^{(g)})$, where $\Phi^*(\cdot)$ is the distribution function of $N[0,1]$, and $f_{m,1}^{(g)}$ and $f_{m,b_m}^{(g)}$ are the frequencies of the first category and the cumulative frequencies of categories with $z_m^{(g)} < b_m$. For analyzing multisample models with interest in group comparisons, it is important to impose conditions for identifying the ordered categorical variables such that the underlying latent continuous variables have the same scale among the groups. To achieve this, we can select the first group as the reference group, and identify its ordered categorical variables by fixing both end thresholds as above. Then, for any m, and $g \neq 1$, we impose the following restrictions (see Lee, Poon and Bentler, 1989),

$$\alpha_{m,k}^{(g)} = \alpha_{m,k}^{(1)}, \qquad k = 1, \cdots, b_m, \tag{6.23}$$

on the thresholds for every ordered categorical variable $z_m^{(g)}$. Under these identification conditions, the unknown parameters in the groups are interpreted in a relative sense, compared over groups. Note that when a different reference group is used, relations over groups are unchanged. Hence, the statistical inferences are unaffected by the choice of the reference group. Clearly, the compatibility of the groups is reflected by the differences of the parameter estimates.

Let $\boldsymbol{\theta}^{(g)}$ be the unknown parameter vector in an identified model and let $\boldsymbol{\alpha}^{(g)}$ be the vector of unknown thresholds that correspond to the gth group. In multisample analysis, a certain type of parameter in $\boldsymbol{\theta}^{(g)}$ is often hypothesized to be invariant over the group models. For example, restrictions on the thresholds, and the following constraints $\boldsymbol{\Lambda}^{(1)} = \cdots = \boldsymbol{\Lambda}^{(G)}$, $\boldsymbol{\Phi}^{(1)} = \cdots = \boldsymbol{\Phi}^{(G)}$, and/or $\boldsymbol{\Gamma}^{(1)} = \cdots = \boldsymbol{\Gamma}^{(G)}$, are often imposed. Let $\boldsymbol{\theta}$ be the vector that contains all unknown distinct parameters in $\boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(G)}$, and let $\boldsymbol{\alpha}$ be the vector that contains all the unknown thresholds. Moreover, let $\mathbf{X}^{(g)} = (\mathbf{x}_1^{(g)}, \cdots, \mathbf{x}_{N_g}^{(g)})$ and $\mathbf{X} = (\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(G)})$ be the observed continuous data, $\mathbf{Z}^{(g)} = (\mathbf{z}_1^{(g)}, \cdots, \mathbf{z}_{N_g}^{(g)})$ and $\mathbf{Z} = (\mathbf{Z}^{(1)}, \cdots, \mathbf{Z}^{(G)})$ be the observed ordered categorical data; let $\mathbf{Y}^{(g)} = (\mathbf{y}_1^{(g)}, \cdots, \mathbf{y}_{N_g}^{(g)})$ and $\mathbf{Y} = (\mathbf{Y}^{(1)}, \cdots, \mathbf{Y}^{(G)})$ be the latent continuous measurement associated with $\mathbf{Z}^{(g)}$ and $\mathbf{Z}$, respectively. Finally, let $\boldsymbol{\Omega}^{(g)} = (\boldsymbol{\omega}_1^{(g)}, \cdots, \boldsymbol{\omega}_{N_g}^{(g)})$ and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}^{(1)}, \cdots, \boldsymbol{\Omega}^{(G)})$ be the matrices of latent variables. In the posterior analysis, the observed data $(\mathbf{X}, \mathbf{Z})$ will be augmented with $(\mathbf{Y}, \boldsymbol{\Omega})$. The Gibbs sampler (Geman and Geman, 1984) will be used to generate a sequence of observations from the joint posterior distribution $[\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}|\mathbf{X}, \mathbf{Z}]$. In applying the Gibbs sampler, we iteratively sample observations from the following conditional distributions: $[\boldsymbol{\Omega}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}]$, $[\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z}]$, and $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z}]$. As in our previous treatments of the thresholds, we assign a noninformative prior to $\boldsymbol{\alpha}$, so that the corresponding prior distribution is proportional to a constant. The first two conditional distributions, which can be derived by the similar reasoning as in Section 6.2 (Appendix 9.1, Lee, 2007). The conditional distribution $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z}]$ is further decomposed into components involving various structural parameters in the different group models. These components are different under various hypotheses of interest or various competing models. Some examples of nonnested competing models (or hypotheses) are:

$$M_A: \quad \text{No constraints}, \qquad M_1: \ \boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(G)}, \quad M_2: \ \boldsymbol{\Lambda}^{(1)} = \cdots = \boldsymbol{\Lambda}^{(G)},$$

$$M_3: \quad \mathbf{\Lambda}_\omega^{(1)} = \cdots = \mathbf{\Lambda}_\omega^{(G)}, \quad M_4: \ \mathbf{\Phi}^{(1)} = \cdots = \mathbf{\Phi}^{(G)}, \tag{6.24}$$
$$M_5: \quad \mathbf{\Psi}_\epsilon^{(1)} = \cdots = \mathbf{\Psi}_\epsilon^{(G)}, \quad M_6: \ \mathbf{\Psi}_\delta^{(1)} = \cdots = \mathbf{\Psi}_\delta^{(G)}.$$

The components in the conditional distribution $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \mathbf{\Omega}, \mathbf{X}, \mathbf{Z}]$ and the specification of prior distributions are slightly different under different M_k defined above. First, prior distributions for nonconstrained parameters in different groups are naturally assumed to be independent. In estimating the unconstrained parameters, we need to specify its own prior distribution, and the data in the corresponding group are used. For constrained parameters across groups, only one prior distribution for these constrained parameters is needed, and all the data in the groups should be combined in the estimation; see Song and Lee (2001). Under this situation, we may not wish to take a joint prior distribution for the factor loading matrix and the unique variance of the error measurement. In the gth group model, let $\psi_{\epsilon k}^{(g)}$ and $\mathbf{\Lambda}_k^{(g)^T}$ are the kth diagonal element of $\mathbf{\Psi}_\epsilon^{(g)}$ and the kth row of $\mathbf{\Lambda}^{(g)}$, respectively. In the multisample analysis, if those parameters are not invariant over groups, their joint prior distribution could be taken as: $\psi_{\epsilon k}^{(g)-1} \stackrel{D}{=} Gamma[\alpha_{0\epsilon k}^{(g)}, \beta_{0\epsilon k}^{(g)}]$, $[\mathbf{\Lambda}_k^{(g)}|\psi_{\epsilon k}^{(g)}] \stackrel{D}{=} N[\mathbf{\Lambda}_{0k}^{(g)}, \psi_{\epsilon k}^{(g)}\mathbf{H}_{0yk}^{(g)}]$, where $\alpha_{0\epsilon k}^{(g)}$, $\beta_{0\epsilon k}^{(g)}$, $\mathbf{\Lambda}_{0k}^{(g)}$, and $\mathbf{H}_{0yk}^{(g)}$ are hyperparameters. This kind of joint prior distribution may cause problems under the constrained situation where $\mathbf{\Lambda}^{(1)} = \cdots = \mathbf{\Lambda}^{(G)} = \mathbf{\Lambda}$, and $\mathbf{\Psi}_\epsilon^{(1)} \neq \cdots \neq \mathbf{\Psi}_\epsilon^{(G)}$, because it is difficult to select $p(\mathbf{\Lambda}_k|\cdot)$ based on a set of different $\psi_{\epsilon k}^{(g)}$. Hence, for convenience, and following the suggestion of Song and Lee (2001), we select independent prior distributions for $\mathbf{\Lambda}^{(g)}$ and $\mathbf{\Psi}_\epsilon^{(g)}$, such that $p(\mathbf{\Lambda}^{(g)}, \mathbf{\Psi}_\epsilon^{(g)}) = p(\mathbf{\Lambda}^{(g)})p(\mathbf{\Psi}_\epsilon^{(g)})$, for $g = 1, 2, \cdots, G$. Under this choice, the prior distribution of $\mathbf{\Lambda}$ under the constraint $\mathbf{\Lambda}^{(1)} = \cdots = \mathbf{\Lambda}^{(G)} = \mathbf{\Lambda}$ is given by: $\mathbf{\Lambda}_k \stackrel{D}{=} N[\mathbf{\Lambda}_{0k}, \mathbf{H}_{0yk}]$, which is independent of $\psi_{\epsilon k}^{(g)}$. Here, $\mathbf{\Lambda}_k^T$ is the kth row of $\mathbf{\Lambda}$. Under the situation $\mathbf{\Lambda}^{(1)} \neq \cdots \neq \mathbf{\Lambda}^{(G)}$, the prior distribution of each $\mathbf{\Lambda}_k^{(g)}$ is given by $N[\mathbf{\Lambda}_{0k}^{(g)}, \mathbf{H}_{0yk}^{(g)}]$. The prior distribution of $\psi_{\epsilon k}^{(g)-1}$ is again $Gamma[\alpha_{0\epsilon k}^{(g)}, \beta_{0\epsilon k}^{(g)}]$. Similarly, we select the prior distributions for $\mathbf{\Lambda}_\omega^{(g)}$ and $\mathbf{\Psi}_\delta^{(g)}$ such that $p(\mathbf{\Lambda}_\omega^{(g)}, \mathbf{\Psi}_\delta^{(g)}) = p(\mathbf{\Lambda}_\omega^{(g)})p(\mathbf{\Psi}_\delta^{(g)})$, for $g = 1, \cdots, G$. Let $\mathbf{\Lambda}_{\omega k}^T$ and $\mathbf{\Lambda}_{\omega k}^{(g)T}$ be the kth rows of $\mathbf{\Lambda}_\omega$ and $\mathbf{\Lambda}_\omega^{(g)}$, respectively. The prior distributions of $\mathbf{\Lambda}_\omega$ are given by:

(i) If $\mathbf{\Lambda}_\omega^{(1)} = \cdots = \mathbf{\Lambda}_\omega^{(G)} = \mathbf{\Lambda}_\omega$, $\quad \mathbf{\Lambda}_{\omega k} \stackrel{D}{=} N[\mathbf{\Lambda}_{0\omega k}, \mathbf{H}_{0\omega k}]$,

(ii) If $\mathbf{\Lambda}_\omega^{(1)} \neq \cdots \neq \mathbf{\Lambda}_\omega^{(G)}$, $\quad \mathbf{\Lambda}_{\omega k}^{(g)} \stackrel{D}{=} N[\mathbf{\Lambda}_{0\omega k}^{(g)}, \mathbf{H}_{0\omega k}^{(g)}]$,

with the hyperparameters $\mathbf{\Lambda}_{0\omega k}$, $\mathbf{\Lambda}_{0\omega k}^{(g)}$, $\mathbf{H}_{0\omega k}$, and $\mathbf{H}_{0\omega k}^{(g)}$. Let $\psi_{\delta k}^{(g)}$ be the kth diagonal element of $\mathbf{\Psi}_\delta^{(g)}$. The prior distribution of $\psi_{\delta k}^{(g)-1}$ is $Gamma[\alpha_{0\delta k}^{(g)}, \beta_{0\delta k}^{(g)}]$, with hyperparamters $\alpha_{0\delta k}^{(g)}$ and $\beta_{0\delta k}^{(g)}$. The prior distributions of $\boldsymbol{\mu}^{(g)}$ and $\mathbf{\Phi}^{(g)}$ are given by:

$$\boldsymbol{\mu}^{(g)} \stackrel{D}{=} N[\boldsymbol{\mu}_0^{(g)}, \mathbf{\Sigma}_0^{(g)}], \quad \mathbf{\Phi}^{(g)-1} \stackrel{D}{=} W_{q_2}[\mathbf{R}_0^{(g)}, \rho_0^{(g)}], \qquad g = 1, \cdots, G,$$

where $\boldsymbol{\mu}_0^{(g)}$, $\mathbf{\Sigma}_0^{(g)}$, $\mathbf{R}_0^{(g)}$, and $\rho_0^{(g)}$ are hyperparameters. Similar adjustments are taken under various combinations of constraints. Based on the above understanding, and the reasoning given in Section 6.2, the conditional distribution $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \mathbf{\Omega}, \mathbf{X}, \mathbf{Z}]$ under various competing models can be obtained. Some results are given in Appendix 6.5.

In the analysis of multisample SEMs, one important statistical inference is on testing whether some types of parameters are invariant over the groups. In Bayesian analysis, each hypothesis of interest is associated with a model, and the problem is approached through model comparison. In contrast to the traditional approach using the likelihood ratio test, it is not necessary to proceed Bayesian model comparison with a sequence of hierarchical models. For instance, depending on the interest of a substantive problem, we can compare any two non-nested models M_k and M_h, as given in (6.24); or compare any M_k with any combination of the models given in (6.24). Similarly, M_k and M_h can be compared using the following Bayes factor:

$$B_{kh} = \frac{p(\mathbf{X}, \mathbf{Z}|M_k)}{p(\mathbf{X}, \mathbf{Z}|M_h)},$$

where $(\mathbf{X}, \mathbf{Z})$ is the observed data set. Let t be a path in $[0, 1]$ to link M_h and M_k, and $0 = t_{(0)} < t_{(1)} < \cdots < t_{(S)} < t_{(S+1)} = 1$ be fixed grids in $[0, 1]$. Let $p(\mathbf{Y}, \mathbf{\Omega}, \mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\alpha}, t)$

be the complete-data likelihood, $\boldsymbol{\theta}$ is the vector of unknown parameters in the link model, and

$$U(\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z}, t) = d \log p(\mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\alpha}, t)/dt.$$

Then,

$$\widehat{\log B_{kh}} = \frac{1}{2}\sum_{s=0}^{S}(t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_{(s)}),$$

where

$$\bar{U}_{(s)} = \frac{1}{J}\sum_{j=1}^{J} U(\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{X}, \mathbf{Z}, t_{(s)}),$$

in which $\{\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \boldsymbol{\Omega}^{(j)} : \ j = 1, 2, \cdots, J\}$ is a sample of observations simulated from the joint posterior distribution $[\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}|\mathbf{X}, \mathbf{Z}, t_{(s)}]$.

To find a good path to link the competing models M_k and M_h is a crucial step in the path sampling procedure for computing $\log B_{kh}$. As an illustrative example, suppose that the competing models M_1 and M_2 are defined as follows. For $g = 1,\ 2$, $i = 1, \cdots, N_g$,

$$\begin{aligned} M_1 : \quad \mathbf{v}_i^{(g)} &= \boldsymbol{\mu}^{(g)} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \\ \boldsymbol{\eta}_i^{(g)} &= \boldsymbol{\Gamma}^{(g)}\mathbf{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\delta}_i^{(g)}; \\ M_2 : \quad \mathbf{v}_i^{(g)} &= \boldsymbol{\mu}^{(g)} + \boldsymbol{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \\ \boldsymbol{\eta}_i^{(g)} &= \boldsymbol{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\delta}_i^{(g)}. \end{aligned}$$

In M_1, $\boldsymbol{\Lambda}$ is invariant over the two groups, while in M_2, $\boldsymbol{\Gamma}$ is invariant over the groups. Note that in both models, the form of the nonlinear terms is the same. Due to the constraints imposed on the parameters, it is rather difficult to find a path t in $[0, 1]$ that directly links M_1 and M_2. This difficulty can be solved through the use of the following auxiliary model M_a:

$$M_a : \quad \mathbf{v}_i^{(g)} = \boldsymbol{\mu}^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \quad g = 1, 2,\ i = 1, \cdots, N_g.$$

The link model M_{ta} for linking M_1 and M_a is define below: For $g = 1, 2, i = 1, \cdots, N_g$,

$$M_{ta1} : \quad \mathbf{v}_i^{(g)} = \boldsymbol{\mu}^{(g)} + t\boldsymbol{\Lambda}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)},$$

$$\boldsymbol{\eta}_i^{(g)} = \boldsymbol{\Gamma}^{(g)}\mathbf{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\delta}_i^{(g)}.$$

When $t = 1$, M_{ta1} reduces to M_1, and when $t = 0$, M_{ta1} reduces to M_a. The parameter vector $\boldsymbol{\theta}$ in M_{ta1} contains $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}_\epsilon^{(1)}$, $\boldsymbol{\Psi}_\epsilon^{(2)}$, $\boldsymbol{\Gamma}^{(1)}$, $\boldsymbol{\Gamma}^{(2)}$, $\boldsymbol{\Phi}^{(1)}$, $\boldsymbol{\Phi}^{(2)}$, $\boldsymbol{\Psi}_\delta^{(1)}$, and $\boldsymbol{\Psi}_\delta^{(2)}$. The link model M_{ta2} for linking M_2 and M_a is defined as follows:

$$\begin{aligned} M_{ta2}: \quad \mathbf{v}_i^{(g)} &= \boldsymbol{\mu}^{(g)} + t\boldsymbol{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \\ \boldsymbol{\eta}_i^{(g)} &= \boldsymbol{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\delta}_i^{(g)}. \end{aligned}$$

Clearly, when $t = 1$ and 0, M_{ta2} reduces to M_2 and M_a, respectively. The parameter vector in M_{ta2} contains $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$, $\boldsymbol{\Lambda}^{(1)}$, $\boldsymbol{\Lambda}^{(2)}$, $\boldsymbol{\Psi}_\epsilon^{(1)}$, $\boldsymbol{\Psi}_\epsilon^{(2)}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}^{(1)}$, $\boldsymbol{\Phi}^{(2)}$, $\boldsymbol{\Psi}_\delta^{(1)}$, and $\boldsymbol{\Psi}_\delta^{(2)}$. We first compute $\log B_{1a}$ and $\log B_{2a}$, and then obtain $\log B_{12}$ via the following equation:

$$\log B_{12} = \log \frac{p(\mathbf{X}, \mathbf{Z}|M_1)/p(\mathbf{X}, \mathbf{Z}|M_a)}{p(\mathbf{X}, \mathbf{Z}|M_2)/p(\mathbf{X}, \mathbf{Z}|M_a)} = \log B_{1a} - \log B_{2a}.$$

Model comparison can also be conducted with DIC through the use of WinBUGS. See the illustrative example in Section 6.3.2.

6.3.2 Analysis of Multisample Quality of Life Data via WinBUGS

Analysis of single group quality of life (QOL) data has been considered in Chapter 5, Section 5.2.4. Here, we describe the Bayesian methods in analyzing multisample QOL data. The WHOQOL-BREF (Power *et al.*, 1999) instrument was taken from the WHOQOL-100 instrument by selecting twenty-six ordered categorical items out of 100 original items. The observations were taken from 15 international field centers, one of which is China, and the rest are western countries, such as the United Kingdom, Italy, and Germany. The first two items are the overall QOL and general health, the next seven items measure physical health, the next six items measure psychological health, the three items that follow are for social relationships, and the last eight items measure the environment. All of the items are measured with a 5-point scale (1 = 'not at all/very dissatisfied'; 2 = 'a little/dissatisfied'; 3 = 'moderate/neither'; 4 = 'very much/satisfied'; and 5 = 'extremely/very satisfied'). To illustrate the Bayesian methodology, we use a synthetic two-sample data that mimic the QOL study with the same items as mentioned above for each sample (see Lee *et al.*, 2005; Lee, 2007). That is, we consider a two-sample SEM on the basis of a simulated data set of randomly drawn observations from two populations. The sample sizes are $N_1 = 338$ and $N_2 = 247$. In the Bayesian analysis, we identify the ordered categorical variables by the method described in Section 6.3.1, using the first group ($g = 1$) as the reference group. Based on the meaning of the questions, we use the following non-overlapping $\mathbf{\Lambda}^{(g)}$ for clear interpretation of latent variables: For $g = 1,\ 2,$

$$\mathbf{\Lambda}^{(g)T} =$$

$$\begin{bmatrix} 1 & \lambda_{2,1}^{(g)} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \lambda_{4,2}^{(g)} & \cdots & \lambda_{9,2}^{(g)} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_{11,3}^{(g)} & \cdots & \lambda_{15,3}^{(g)} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_{17,4}^{(g)} & \lambda_{18,4}^{(g)} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 1 & \lambda_{20,5}^{(g)} & \cdots & \lambda_{26,5}^{(g)} \end{bmatrix},$$

where 1's and 0's are fixed parameters. The latent variables in $\boldsymbol{\omega}_i^{(g)T} = (\eta_i^{(g)}, \xi_{i1}^{(g)}, \xi_{i2}^{(g)}, \xi_{i3}^{(g)}, \xi_{i4}^{(g)})$ are interpreted as 'health related QOL, η', 'physical health, ξ_1', 'psychological health, ξ_2', 'social relationship, ξ_3', and 'environment, ξ_4'. The measurement equation in the model is given by

$$\mathbf{v}_i^{(g)} = \boldsymbol{\mu}^{(g)} + \mathbf{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \qquad g = 1, 2,$$

with $\mathbf{\Lambda}^{(g)}$ defined as above. The following structural equation is used to assess the effects of the latent constructs in $\boldsymbol{\xi}_i^{(g)}$ on the health related QOL, $\eta_i^{(g)}$:

$$\eta_i^{(g)} = \gamma_1^{(g)}\xi_{i1}^{(g)} + \gamma_2^{(g)}\xi_{i2}^{(g)} + \gamma_3^{(g)}\xi_{i3}^{(g)} + \gamma_4^{(g)}\xi_{i4}^{(g)} + \delta_i^{(g)}.$$

In the Bayesian analysis, the prior inputs of the hyperparameters in the conjugate prior distributions are taken as follows: $\alpha_{0\epsilon k}^{(g)} = \alpha_{0\delta k}^{(g)} = 10$, $\beta_{0\epsilon k}^{(g)} = \beta_{0\delta k}^{(g)} = 8$, elements in $\mathbf{\Lambda}_{0k}^{(g)}$ are taken as 0.8, elements in $\mathbf{\Lambda}_{0\omega k}^{(g)}$ are taken as 0.6, $\mathbf{H}_{0yk}^{(g)}$ and $\mathbf{H}_{0\omega k}^{(g)}$ are diagonal matrices with diagonal elements 0.25, $\mathbf{R}_0^{(g)^{-1}} = 8\mathbf{I}_4$, and $\rho_0^{(g)} = 30$. Three multisample models M_1, M_2, and M_3 that are respectively associated with following hypotheses are considered: H_1: No constraints; $H_2: \ \mathbf{\Lambda}^{(1)} = \mathbf{\Lambda}^{(2)}$; and $H_3: \ \mathbf{\Lambda}^{(1)} = \mathbf{\Lambda}^{(2)}, \mathbf{\Phi}^{(1)} = \mathbf{\Phi}^{(2)}$. The software WinBUGS was applied to obtain the Bayesian results. In the analysis, the number of burn-in iterations was taken as 10,000, and additional 10,000 observations were collected to produce the results. The DIC values corresponding to M_1, M_2, and M_3 are equal to 32302.6, 32321.7, and 32341.9, respectively. Hence, model M_1 with the smallest DIC value is selected. The Bayesian estimates and their standard error estimates produced by WinBUGS under M_1 are presented in Table 6.4. The WinBUGS codes in analyzing models M_1, M_2, and M_3 are given in the following websites:

http://www.sta.cuhk.edu.hk/song-lee/book-chapter6(section6.3.2)/WinBUGS-M1-code

http://www.sta.cuhk.edu.hk/song-lee/book-chapter6(section6.3.2)/WinBUGS-M2-code

http://www.sta.cuhk.edu.hk/song-lee/book-chapter6(section6.3.2)/WinBUGS-M3-code

http://www.sta.cuhk.edu.hk/song-lee/book-chapter6(section6.3.2)/WinBUGS-data.

(PLEASE CHANGE TO THE WEB-SITES HOUSED IN JOHN-WILEY).

Table 6.4 here

Appendix 6.1: Conditional Distributions: Two-level Nonlinear SEM

Owing to the complexity of the model, it is very tedious to derive all the conditional distributions required by the Gibbs sampler, hence only brief discussions are given. For brevity, we will use $p(\cdot|\cdot)$ to denote the conditional distribution if the context is clear. Moreover, we only consider the case that all parameters in $\mathbf{\Lambda}_{1g}$, $\mathbf{\Lambda}_2$, $\mathbf{\Pi}_{1g}$, $\mathbf{\Gamma}_{1g}$, $\mathbf{\Pi}_2$, and $\mathbf{\Gamma}_2$ are not fixed. Conditional distributions for the case with fixed parameters can be obtained by slight modifications as given in previous chapters.

$p(\mathbf{V}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{\Omega}_1, \mathbf{\Omega}_2, \mathbf{X}, \mathbf{Z})$: Since $\mathbf{v}_g$'s are independent and not depending on $\boldsymbol{\alpha}$, this conditional distribution is equal to a product of $p(\mathbf{v}_g|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}_g, \mathbf{\Omega}_{1g}, \boldsymbol{\omega}_{2g}, \mathbf{X}_g, \mathbf{Z}_g)$ with $g = 1, \cdots, G$. For each gth term in this product,

$$p(\mathbf{v}_g|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}_g, \mathbf{\Omega}_{1g}, \boldsymbol{\omega}_{2g}, \mathbf{X}_g, \mathbf{Z}_g) \propto p(\mathbf{v}_g|\boldsymbol{\theta}, \boldsymbol{\omega}_{2g}) \prod_{i=1}^{N_g} p(\mathbf{u}_{gi}|\boldsymbol{\theta}, \mathbf{v}_g, \boldsymbol{\omega}_{1gi}) \tag{6.A1}$$

$$\propto \exp\left[-\frac{1}{2}\left\{\mathbf{v}_g^T(N_g\mathbf{\Psi}_{1g}^{-1} + \mathbf{\Psi}_2^{-1})\mathbf{v}_g - 2\mathbf{v}_g^T\left[\mathbf{\Psi}_{1g}^{-1}\sum_{i=1}^{N_g}(\mathbf{u}_{gi} - \mathbf{\Lambda}_{1g}\boldsymbol{\omega}_{1gi}) + \mathbf{\Psi}_2^{-1}(\boldsymbol{\mu} + \mathbf{\Lambda}_2\boldsymbol{\omega}_{2g})\right]\right\}\right]$$

Hence, for each $\mathbf{v}_g$, its conditional distribution $p(\mathbf{v}_g|\cdot)$ is $N[\boldsymbol{\mu}_g^*, \mathbf{\Sigma}_g^*]$, where

$$\boldsymbol{\mu}_g^* = \mathbf{\Sigma}_g^*[\mathbf{\Psi}_{1g}^{-1}\sum_{i=1}^{N_g}(\mathbf{u}_{gi} - \mathbf{\Lambda}_{1g}\boldsymbol{\omega}_{1gi}) + \mathbf{\Psi}_2^{-1}(\boldsymbol{\mu} + \mathbf{\Lambda}_2\boldsymbol{\omega}_{2g})], \quad \text{and} \quad \mathbf{\Sigma}_g^* = (N_g\mathbf{\Psi}_{1g}^{-1} + \mathbf{\Psi}_2^{-1})^{-1}.$$

$p(\mathbf{\Omega}_1|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \mathbf{\Omega}_2, \mathbf{X}, \mathbf{Z})$: Since $\boldsymbol{\omega}_{1gi}$ are mutually independent, $\mathbf{u}_{gi}$ is independent with $\mathbf{u}_{hi}$ for all $h \neq g$, and they are not depending on $\boldsymbol{\alpha}$ and $\mathbf{Y}$, we have

$$p(\mathbf{\Omega}_1|\cdot) = \prod_{g=1}^{G}\prod_{i=1}^{N_g} p(\boldsymbol{\omega}_{1gi}|\boldsymbol{\theta}, \mathbf{v}_g, \boldsymbol{\omega}_{2g}, \mathbf{u}_{gi}) \propto \prod_{g=1}^{G}\prod_{i=1}^{N_g} p(\mathbf{u}_{gi}|\boldsymbol{\theta}, \mathbf{v}_g, \boldsymbol{\omega}_{1gi})\, p(\boldsymbol{\eta}_{1gi}|\boldsymbol{\xi}_{1gi}, \boldsymbol{\theta})p(\boldsymbol{\xi}_{1gi}|\boldsymbol{\theta}).$$

It follows that $p(\boldsymbol{\omega}_{1gi}|\cdot)$ is proportional to

$$\begin{aligned}\exp\Big[&-\frac{1}{2}\{\boldsymbol{\xi}_{1gi}^T\mathbf{\Phi}_{1g}^{-1}\boldsymbol{\xi}_{1gi} + (\mathbf{u}_{gi} - \mathbf{v}_g - \mathbf{\Lambda}_{1g}\boldsymbol{\omega}_{1gi})^T\mathbf{\Psi}_{1g}^{-1}(\mathbf{u}_{gi} - \mathbf{v}_g - \mathbf{\Lambda}_{1g}\boldsymbol{\omega}_{1gi}) \\ &+ [\boldsymbol{\eta}_{1gi} - \mathbf{\Pi}_{1g}\boldsymbol{\eta}_{1gi} - \mathbf{\Gamma}_{1g}\mathbf{F}_1(\boldsymbol{\xi}_{1gi})]^T\mathbf{\Psi}_{1g\delta}^{-1}[\boldsymbol{\eta}_{1gi} - \mathbf{\Pi}_{1g}\boldsymbol{\eta}_{1gi} - \mathbf{\Gamma}_{1g}\mathbf{F}_1(\boldsymbol{\xi}_{1gi})]\}\Big].\end{aligned} \tag{6.A2}$$

$p(\mathbf{\Omega}_2|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \mathbf{\Omega}_1, \mathbf{X}, \mathbf{Z})$: This distribution has very similar form as in $p(\mathbf{\Omega}_1|\cdot)$ and (6.A2), hence it is not presented.

$p(\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z})$: We only consider the case that all the thresholds corresponding to each within-group are different. The other cases can be similarly derived. To deal with the situation with little or no information about these parameters, the following noninformative prior distribution is used:

$$p(\alpha_{gk}) = p(\alpha_{gk,2}, \cdots, \alpha_{gk,b_k-1}) \propto C, \quad g = 1, \cdots, G, \quad k = 1, \cdots, s,$$

where C is a constant. Now, since $(\boldsymbol{\alpha}_g, \mathbf{Y}_g)$ is independent with $(\boldsymbol{\alpha}_h, \mathbf{Y}_h)$ for $g \neq h$, and that $\boldsymbol{\Psi}_{1g}$ is diagonal,

$$p(\boldsymbol{\alpha}, \mathbf{Y}|\cdot) = \prod_{g=1}^{G} p(\boldsymbol{\alpha}_g, \mathbf{Y}_g|\cdot) = \prod_{g=1}^{G} \prod_{k=1}^{s} p(\boldsymbol{\alpha}_{gk}, \mathbf{Y}_{gk}|\cdot), \tag{6.A3}$$

where $\mathbf{Y}_{gk} = (y_{g1k}, \cdots, y_{gN_gk})^T$. Let $\boldsymbol{\Psi}_{1gy}$, $\boldsymbol{\Lambda}_{1gy}$, and $\mathbf{v}_{gy}$ be the submatrices and subvector of $\boldsymbol{\Psi}_{1g}$, $\boldsymbol{\Lambda}_{1g}$, and $\mathbf{v}_g$ corresponding to the ordered categorical variables in $\mathbf{Y}$; let ψ_{1gyk} be the kth diagonal element of $\boldsymbol{\Psi}_{1gy}$, $\boldsymbol{\Lambda}_{1gyk}^T$ be the kth row of $\boldsymbol{\Lambda}_{1gy}$, v_{gyk} be the kth element of $\mathbf{v}_{gy}$, and $I_A(y)$ be an indicator function with value 1 if y in A and zero otherwise, $p(\boldsymbol{\alpha}, \mathbf{Y}|\cdot)$ can be obtained from (6.A3) and

$$p(\boldsymbol{\alpha}_{gk}, \mathbf{Y}_{gk}|\cdot) \propto \prod_{i=1}^{N_g} \phi\{\psi_{1gyk}^{-1/2}(y_{gik} - v_{gyk} - \boldsymbol{\Lambda}_{1gyk}^T \boldsymbol{\omega}_{1gi})\} I_{(\alpha_{gk,z_{gik}}, \alpha_{gk,z_{gik}+1}]}(y_{gik}), \tag{6.A4}$$

where ϕ is the probability density function of $N[0, 1]$.

$p(\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z})$: This conditional distribution is different under different special cases as discussed in Section 6.2.2. We first consider the situation with distinct within-group parameters, that is $\boldsymbol{\theta}_{11} \neq \cdots \neq \boldsymbol{\theta}_{1G}$. Let $\boldsymbol{\theta}_2$ be the vector of unknown parameters in $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}_2$, and $\boldsymbol{\Psi}_2$; and $\boldsymbol{\theta}_{2\omega}$ be the vector of unknown parameters in $\boldsymbol{\Pi}_2$, $\boldsymbol{\Gamma}_2$, $\boldsymbol{\Phi}_2$, and $\boldsymbol{\Psi}_{2\delta}$. These between-group parameters are the same for each g. For the within-group parameters, let $\boldsymbol{\theta}_{1g}$ be the vector of unknown parameters in $\boldsymbol{\Lambda}_{1g}$ and $\boldsymbol{\Psi}_{1g}$; and $\boldsymbol{\theta}_{1g\omega}$ be the vector of unknown parameters in $\boldsymbol{\Pi}_{1g}$, $\boldsymbol{\Gamma}_{1g}$, $\boldsymbol{\Phi}_{1g}$, and $\boldsymbol{\Psi}_{1g\delta}$. It is natural to assume the prior distributions of these parameter vectors in different independent groups are independent to each other, and hence they can be treated separately.

For $\boldsymbol{\theta}_{1g}$, the following commonly used conjugate type prior distributions are used:

$$\psi_{1gk}^{-1} \stackrel{D}{=} Gamma[\alpha_{01gk}, \beta_{01gk}], \quad [\boldsymbol{\Lambda}_{1gk}|\psi_{1gk}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{01gk}, \psi_{1gk}\mathbf{H}_{01gk}], \quad k = 1, \cdots, p,$$

where ψ_{1gk} is the kth diagonal element of $\boldsymbol{\Psi}_{1g}$, $\boldsymbol{\Lambda}_{1gk}^T$ is the kth row of $\boldsymbol{\Lambda}_{1g}$, and $\alpha_{01gk}, \beta_{01gk}, \boldsymbol{\Lambda}_{01gk}$, and $\mathbf{H}_{01gk}$ are given hyperparameter values. For $k \neq h$, it is assumed that $(\psi_{1gk}, \boldsymbol{\Lambda}_{1gk})$ and $(\psi_{1gh}, \boldsymbol{\Lambda}_{1gh})$ are independent. Let $\mathbf{U}_g^* = \{\mathbf{u}_{gi} - \mathbf{v}_g,\ i = 1, \cdots, N_g\}$ and $\mathbf{U}_{gk}^{*T}$ be the kth row of $\mathbf{U}_g^*$, $\boldsymbol{\Sigma}_{1gk} = (\mathbf{H}_{01gk}^{-1} + \boldsymbol{\Omega}_{1g}\boldsymbol{\Omega}_{1g}^T)^{-1}$, $\mathbf{m}_{1gk} = \boldsymbol{\Sigma}_{1gk}(\mathbf{H}_{01gk}^{-1}\boldsymbol{\Lambda}_{01gk} + \boldsymbol{\Omega}_{1g}\mathbf{U}_{gk}^*)$, $\boldsymbol{\Omega}_{1g} = (\boldsymbol{\omega}_{1g1}, \cdots, \boldsymbol{\omega}_{1gN_g})$, and $\beta_{1gk} = \beta_{01gk} + (\mathbf{U}_{gk}^{*T}\mathbf{U}_{gk}^* - \mathbf{m}_{1gk}^T\boldsymbol{\Sigma}_{1gk}^{-1}\mathbf{m}_{1gk} + \boldsymbol{\Lambda}_{01gk}^T\mathbf{H}_{01gk}^{-1}\boldsymbol{\Lambda}_{01gk})/2$, it can be shown that

$$[\psi_{1gk}^{-1}|\cdot] \stackrel{D}{=} Gamma\ (N_g/2 + \alpha_{01gk}, \beta_{1gk}), \quad [\boldsymbol{\Lambda}_{1gk}|\psi_{1gk}, \cdot] \stackrel{D}{=} N[\mathbf{m}_{1gk}, \psi_{1gk}\boldsymbol{\Sigma}_{1gk}]. \qquad (6.A5)$$

For $\boldsymbol{\theta}_{1g\omega}$, it is assumed that $\boldsymbol{\Phi}_{1g}$ is independent with $(\boldsymbol{\Lambda}_{1g}^*, \boldsymbol{\Psi}_{1g\delta})$, where $\boldsymbol{\Lambda}_{1g}^* = (\boldsymbol{\Pi}_{1g}, \boldsymbol{\Gamma}_{1g})$. Also, $(\boldsymbol{\Lambda}_{1gk}^*, \psi_{1g\delta k})$ and $(\boldsymbol{\Lambda}_{1gh}^*, \psi_{1g\delta h})$ are independent, where $\boldsymbol{\Lambda}_{1gk}^{*T}$ and $\psi_{1g\delta k}$ are the kth row and diagonal element of $\boldsymbol{\Lambda}_{1g}^*$ and $\boldsymbol{\Psi}_{1g\delta}$, respectively. The associated prior distribution of $\boldsymbol{\Phi}_{1g}$ is: $\boldsymbol{\Phi}_{1g}^{-1} \stackrel{D}{=} W_{q_{12}}[\mathbf{R}_{01g}, \rho_{01g}]$, where ρ_{01g} and the positive definite matrix $\mathbf{R}_{01g}$ are given hyperparameters. Moreover, the prior distribution of $\psi_{1g\delta k}$ and $\boldsymbol{\Lambda}_{1gk}^*$ are:

$$\psi_{1g\delta k}^{-1} \stackrel{D}{=} Gamma[\alpha_{01g\delta k}, \beta_{01g\delta k}], \quad [\boldsymbol{\Lambda}_{1gk}^*|\psi_{1g\delta k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{01gk}^*, \psi_{1g\delta k}\mathbf{H}_{01gk}^*], \quad k = 1, \cdots, q_{11},$$

where $\alpha_{01g\delta k}$, $\beta_{01g\delta k}$, $\boldsymbol{\Lambda}_{01gk}^*$, and $\mathbf{H}_{01gk}^*$ are given hyperparameters. Let $\boldsymbol{\Omega}_{1g}^* = \{\boldsymbol{\eta}_{1g1}, \cdots, \boldsymbol{\eta}_{1gN_g}\}$, $\boldsymbol{\Omega}_{1gk}^{*T}$ be the kth row of $\boldsymbol{\Omega}_{1g}^*$, $\boldsymbol{\Xi}_{1g} = \{\boldsymbol{\xi}_{1g1}, \cdots, \boldsymbol{\xi}_{1gN_g}\}$ and $\mathbf{F}_{1g}^* = \{\mathbf{F}_1^*(\boldsymbol{\omega}_{1g1}), \cdots, \mathbf{F}_1^*(\boldsymbol{\omega}_{1gN_g})\}$, in which $\mathbf{F}_1^*(\boldsymbol{\omega}_{1gi}) = (\boldsymbol{\eta}_{1gi}^T, \mathbf{F}_1(\boldsymbol{\xi}_{1gi})^T)^T$, $i = 1, \cdots, N_g$, it can be shown that

$$[\psi_{1g\delta k}^{-1}|\cdot] \stackrel{D}{=} Gamma\ [N_g/2 + \alpha_{01g\delta k}, \beta_{1g\delta k}], \quad [\boldsymbol{\Lambda}_{1gk}^*|\psi_{1g\delta k}^{-1}, \cdot] \stackrel{D}{=} N[\mathbf{m}_{1gk}^*, \psi_{1g\delta k}\boldsymbol{\Sigma}_{1gk}^*], \quad (6.A6)$$

where $\boldsymbol{\Sigma}_{1gk}^* = (\mathbf{H}_{01gk}^{*-1} + \mathbf{F}_{1g}^*\mathbf{F}_{1g}^{*T})^{-1}$, $\mathbf{m}_{1gk}^* = \boldsymbol{\Sigma}_{1gk}^*(\mathbf{H}_{01gk}^{*-1}\boldsymbol{\Lambda}_{01gk}^* + \mathbf{F}_{1g}^*\boldsymbol{\Omega}_{1gk}^*)$, and $\beta_{1g\delta k} = \beta_{01g\delta k} + (\boldsymbol{\Omega}_{1gk}^{*T}\boldsymbol{\Omega}_{1gk}^* - \mathbf{m}_{1gk}^{*T}\boldsymbol{\Sigma}_{1gk}^{*-1}\mathbf{m}_{1gk}^* + \boldsymbol{\Lambda}_{01gk}^{*T}\mathbf{H}_{01gk}^{*-1}\boldsymbol{\Lambda}_{01gk}^*)/2$. The conditional distribution relating to $\boldsymbol{\Phi}_{1g}$ is given by

$$[\boldsymbol{\Phi}_{1g}|\boldsymbol{\Xi}_{1g}] \stackrel{D}{=} IW_{q_{12}}[(\boldsymbol{\Xi}_{1g}\boldsymbol{\Xi}_{1g}^T + \mathbf{R}_{01g}^{-1}), N_g + \rho_{01g}]. \qquad (6.A7)$$

Conditional distributions involved in $\boldsymbol{\theta}_2$ are derived similarly on the basis of the following independent conjugate type prior distributions: for $k = 1, \cdots, p$, and

$$\boldsymbol{\mu} \stackrel{D}{=} N[\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0], \quad \psi_{2k}^{-1} \stackrel{D}{=} Gamma[\alpha_{02k}, \beta_{02k}], \quad [\boldsymbol{\Lambda}_{2k}|\psi_{2k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{02k}, \psi_{2k}\mathbf{H}_{02k}].$$

where $\boldsymbol{\Lambda}_{2k}^T$ is the kth row of $\boldsymbol{\Lambda}_2$, ψ_{2k} is the kth diagonal element of $\boldsymbol{\Psi}_2$, $\alpha_{02k},\ \beta_{02k},\ \boldsymbol{\mu}_0,$ $\boldsymbol{\Sigma}_0,\ \boldsymbol{\Lambda}_{02k}$, and $\mathbf{H}_{02k}$ are given hyperparameters.

Similarly, conditional distributions involved in $\boldsymbol{\theta}_{2\omega}$ are derived on the basis of the following conjugate type distributions: for $k = 1, \cdots, q_{21}$,

$$\boldsymbol{\Phi}_2^{-1} \stackrel{D}{=} W_{q_{22}}[\mathbf{R}_{02}, \rho_{02}], \ \psi_{2\delta k}^{-1} \stackrel{D}{=} Gamma[\alpha_{02\delta k}, \beta_{02\delta k}], \ [\boldsymbol{\Lambda}_{2k}^*|\psi_{2\delta k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{02k}^*, \psi_{2\delta k}\mathbf{H}_{02k}^*],$$

where $\boldsymbol{\Lambda}_2^* = (\boldsymbol{\Pi}_2, \boldsymbol{\Gamma}_2)$, $\boldsymbol{\Lambda}_{2k}^{*T}$ is the vector that contains the unknown parameters in the kth row of $\boldsymbol{\Lambda}_2^*$, and $\psi_{2\delta k}$ is the kth diagonal element of $\boldsymbol{\Psi}_{2\delta}$. As these conditional distributions are similar to those in (6.A5)$-$(6.A7), they are not presented here.

Under the situation that $\boldsymbol{\theta}_{11} = \cdots = \boldsymbol{\theta}_{1G} (= \boldsymbol{\theta}_1)$, the prior distributions corresponding to components of $\boldsymbol{\theta}_1$ are not depending on g, and all the data in the within groups should be combined in deriving the conditional distributions for the estimation. Conditional distributions can be derived with the following conjugate type prior distributions: for $k = 1, \cdots, p$ and similar notations as above,

$$\begin{aligned} &\psi_{1k}^{-1} \stackrel{D}{=} Gamma[\alpha_{01k}, \beta_{01k}], \quad [\boldsymbol{\Lambda}_{1k}|\psi_{1k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{01k}, \psi_{1k}\mathbf{H}_{01k}], \quad \boldsymbol{\Phi}_1^{-1} \stackrel{D}{=} W_{q_{12}}[\mathbf{R}_{01}, \rho_{01}], \\ &\psi_{1\delta k}^{-1} \stackrel{D}{=} Gamma[\alpha_{01\delta k}, \beta_{01\delta k}], \quad [\boldsymbol{\Lambda}_{1k}^*|\psi_{1\delta k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{01k}^*, \psi_{1\delta k}\mathbf{H}_{01k}^*], \end{aligned} \tag{6.A8}$$

and the prior distributions and conditional distributions corresponding to structural parameters in the between-group covariance matrix are the same as before.

Appendix 6.2: The MH Algorithm: Two-level Nonlinear SEM

Simulating observations from the Gamma, normal and inverted Wishart distributions is straightforward and fast. However, the conditional distributions, $p(\boldsymbol{\Omega}_1|\cdot)$, $p(\boldsymbol{\Omega}_2|\cdot)$, and

$p(\boldsymbol{\alpha}, \mathbf{Y}|\cdot)$, are complex, hence it is necessary to implement the MH algorithm for efficient simulation of observations from these conditional distributions.

For the conditional distribution $p(\boldsymbol{\Omega}_1|\cdot)$, we require to simulate observations from the target density $p(\boldsymbol{\omega}_{1gi}|\cdot)$ as given in (6.A2). Similar to the method of Zhu and Lee (1999), and Lee and Zhu (2000), we choose $N[\cdot, \sigma_1^2\mathbf{D}_{1g}]$ as the proposal distribution, where $\mathbf{D}_{1g}^{-1} = \mathbf{D}_{1gw}^{-1} + \boldsymbol{\Lambda}_{1g}^T\boldsymbol{\Psi}_{1g}^{-1}\boldsymbol{\Lambda}_{1g}$, and $\mathbf{D}_{1gw}^{-1}$ is given by

$$\mathbf{D}_{1gw}^{-1} = \begin{bmatrix} \boldsymbol{\Pi}_{1g0}^T\boldsymbol{\Psi}_{1g\delta}^{-1}\boldsymbol{\Pi}_{1g0} & -\boldsymbol{\Pi}_{1g0}^T\boldsymbol{\Psi}_{1g\delta}^{-1}\boldsymbol{\Gamma}_{1g}\boldsymbol{\Delta}_{1g} \\ -\boldsymbol{\Delta}_{1g}^T\boldsymbol{\Gamma}_{1g}^T\boldsymbol{\Psi}_{1g\delta}^{-1}\boldsymbol{\Pi}_{1g0} & \boldsymbol{\Phi}_{1g}^{-1} + \boldsymbol{\Delta}_{1g}^T\boldsymbol{\Gamma}_{1g}^T\boldsymbol{\Psi}_{1g\delta}^{-1}\boldsymbol{\Gamma}_{1g}\boldsymbol{\Delta}_{1g} \end{bmatrix},$$

where $\boldsymbol{\Pi}_{1g0} = \mathbf{I}_{q11} - \boldsymbol{\Pi}_{1g}$ with an identity matrix $\mathbf{I}_{q11}$ of order q_{11}, and $\boldsymbol{\Delta}_{1g} = (\partial\mathbf{H}_1(\boldsymbol{\xi}_{1gi})/\partial\boldsymbol{\xi}_{1gi})^T|_{\boldsymbol{\xi}_{1gi}=\mathbf{0}}$. Let $p(\cdot|\boldsymbol{\omega}^*, \sigma_1^2\mathbf{D}_{1g})$ be the density function corresponding to the proposal distribution $N[\boldsymbol{\omega}^*, \sigma_1^2\mathbf{D}_{1g}]$, the MH algorithm is implemented as follows: At the mth iteration with a current value $\boldsymbol{\omega}_{1gi}^{(m)}$, a new candidate $\boldsymbol{\omega}_i^*$ is generated from $p(\cdot|\boldsymbol{\omega}_{1gi}^{(m)}, \sigma_1^2\mathbf{D}_{1g})$ and accepting this new candidate with probability $\min\{1, p(\boldsymbol{\omega}_i^*|\cdot)/p(\boldsymbol{\omega}_{1gi}^{(m)}|\cdot)\}$. The variance σ_1^2 can be chosen such that the average acceptance rate is approximately 0.25 or more, see Gelman, Roberts and Gilks (1996).

Observations from the conditional distribution $p(\boldsymbol{\Omega}_2|\cdot)$ with target density similar to (6.A2) can be simulated via a similar MH algorithm as described above. To save space, details are not given.

An MH type algorithm is necessary for simulating observations from the complex distribution $p(\boldsymbol{\alpha}, \mathbf{Y}|\cdot)$. Here, the target density is given in (6.A4). According to the factorization recommended by Cowles (1996), see also Lee and Zhu (2000), the joint proposal density of $\boldsymbol{\alpha}_{gk}$ and $\mathbf{Y}_{gk}$ is constructed as

$$p(\boldsymbol{\alpha}_{gk}, \mathbf{Y}_{gk}|\cdot) = p(\boldsymbol{\alpha}_{gk}|\ \cdot)p(\mathbf{Y}_{gk}|\boldsymbol{\alpha}_{gk}, \cdot). \tag{6.A9}$$

Then, the algorithm is implemented as follows: At the mth iteration with $(\boldsymbol{\alpha}_{gk}^{(m)}, \mathbf{Y}_{gk}^{(m)})$, the acceptance probability for a $(\boldsymbol{\alpha}_{gk}, \mathbf{Y}_{gk})$ as a new observation $(\boldsymbol{\alpha}_{gk}^{(m+1)}, \mathbf{Y}_{gk}^{(m+1)})$ is

$\min\{1, R_{gk}\}$, where

$$R_{gk} = \frac{p(\boldsymbol{\alpha}_{gk}, \mathbf{Y}_{gk}|\boldsymbol{\theta}, \mathbf{Z}_{gk}, \boldsymbol{\Omega}_{1g})p(\boldsymbol{\alpha}_{gk}^{(m)}, \mathbf{Y}_{gk}^{(m)}|\boldsymbol{\alpha}_{gk}, \mathbf{Y}_{gk}, \boldsymbol{\theta}, \mathbf{Z}_{gk}, \boldsymbol{\Omega}_{1g})}{p(\boldsymbol{\alpha}_{gk}^{(m)}, \mathbf{Y}_{gk}^{(m)}|\boldsymbol{\theta}, \mathbf{Z}_{gk}, \boldsymbol{\Omega}_{1g})p(\boldsymbol{\alpha}_{gk}, \mathbf{Y}_{gk}|\boldsymbol{\alpha}_{gk}^{(m)}, \mathbf{Y}_{gk}^{(m)}, \boldsymbol{\theta}, \mathbf{Z}_{gk}, \boldsymbol{\Omega}_{1g})}.$$

To search for a new observation via the proposal density (6.A9), we first generate a vector of thresholds $(\alpha_{gk,2}, \cdots, \alpha_{gk,b_k-1})$ from the following truncated normal distribution

$$\alpha_{gk,z} \stackrel{D}{=} N[\alpha_{gk,z}^{(m)}, \sigma_{\alpha gk}^2] I_{(\alpha_{gk,z-1}, \alpha_{gk,z+1}^{(m)}]}(\alpha_{gk,z}), \quad z = 2, \cdots, b_k - 1,$$

where $\sigma_{\alpha gk}^2$ is a preassigned value to give an approximate acceptance rate 0.44, see Cowles (1996). It follows from (6.A4) and the above result that

$$R_{gk} = \prod_{z=2}^{b_k-1} \frac{\Phi^*\{(\alpha_{gk,z+1}^{(m)} - \alpha_{gk,z}^{(m)})/\sigma_{\alpha gk}\} - \Phi^*\{(\alpha_{gk,z-1} - \alpha_{gk,z}^{(m)})/\sigma_{\alpha gk}\}}{\Phi^*\{(\alpha_{gk,z+1} - \alpha_{gk,z})/\sigma_{\alpha gk}\} - \Phi^*\{(\alpha_{gk,z-1}^{(m)} - \alpha_{gk,z})/\sigma_{\alpha gk}\}} \times$$
$$\prod_{i=1}^{N_g} \frac{\Phi^*\{\psi_{1gyk}^{-1/2}(\alpha_{gk,z_{gik}+1} - v_{gyk} - \boldsymbol{\Lambda}_{1gyk}^T \boldsymbol{\omega}_{1gi})\} - \Phi^*\{\psi_{1gyk}^{-1/2}(\alpha_{gk,z_{gik}} - v_{gyk} - \boldsymbol{\Lambda}_{1gyk}^T \boldsymbol{\omega}_{1gi})\}}{\Phi^*\{\psi_{1gyk}^{-1/2}(\alpha_{gk,z_{gik}+1}^{(m)} - v_{gyk} - \boldsymbol{\Lambda}_{1gyk}^T \boldsymbol{\omega}_{1gi})\} - \Phi^*\{\psi_{1gyk}^{-1/2}(\alpha_{gk,z_{gik}}^{(m)} - v_{gyk} - \boldsymbol{\Lambda}_{1gyk}^T \boldsymbol{\omega}_{1gi})\}},$$

where Φ^* is the distribution function of $N[0, 1]$. Since R_{gk} only depends on the old and new values of $\boldsymbol{\alpha}_{gk}$ but not on $\mathbf{Y}_{gk}$, it only requires to generate a new $\mathbf{Y}_{gk}$ for an accepted $\boldsymbol{\alpha}_{gk}$. This new $\mathbf{Y}_{gk}$ is simulated from the truncated normal distribution $p(\mathbf{Y}_{gk}|\boldsymbol{\alpha}_{gk}, \cdot)$ via the algorithm given in Robert (1995).

Appendix 6.3: PP p-value for Two-level Nonlinear SEM with Mixed Continuous and Ordered Categorical Variables

Suppose null hypothesis H_0 is that the model defined in (6.1) and (6.2) is plausible, the PP p-value is defined as

$$p_B(\mathbf{X}, \mathbf{Z}) = Pr\{D(\mathbf{U}^{\text{rep}}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) \geq D(\mathbf{U}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)|\mathbf{X}, \mathbf{Z}, H_0\},$$

where $\mathbf{U}^{\text{rep}}$ denotes a replication of $\mathbf{U} = \{\mathbf{u}_{gi},\ i = 1, \cdots, N_g,\ g = 1, \cdots, G\}$, with $\mathbf{u}_{gi}$ satisfies the model defined by equation (6.3) that involves structural parameters and latent variables satisfying (6.4) and (6.5), and $D(\cdot|\cdot)$ is a discrepancy variable. Here, the following χ^2 discrepancy variable is used:

$$D(\mathbf{U}^{\text{rep}}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = \sum_{g=1}^{G}\sum_{i=1}^{N_g}(\mathbf{u}_{gi}^{\text{rep}} - \mathbf{v}_g - \boldsymbol{\Lambda}_{1g}\boldsymbol{\omega}_{1gi})^T\boldsymbol{\Psi}_{1g}^{-1}(\mathbf{u}_{gi}^{\text{rep}} - \mathbf{v}_g - \boldsymbol{\Lambda}_{1g}\boldsymbol{\omega}_{1gi}),$$

which is distributed as $\chi^2(pn)$, a χ^2 distribution with pn degrees of freedom. Here, $n = N_1 + \cdots + N_g$. The PP p-value on the basis of this discrepancy variable is

$$P_B(\mathbf{X}, \mathbf{Z}) = \int Pr\{\chi^2(pn) \geq D(\mathbf{U}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)\}\times$$
$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2|\mathbf{X}, \mathbf{Z})d\boldsymbol{\theta}d\boldsymbol{\alpha}d\mathbf{Y}d\mathbf{V}d\boldsymbol{\Omega}_1 d\boldsymbol{\Omega}_2.$$

A Rao-Blackwellized type estimate of the PP p-value is equal to

$$\hat{P}_B(\mathbf{X}, \mathbf{Z}) = T^{-1}\sum_{t=1}^{T} Pr\{\chi^2(pn) \geq D(\mathbf{U}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\alpha}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{V}^{(t)}, \boldsymbol{\Omega}_1^{(t)}, \boldsymbol{\Omega}_2^{(t)})\}$$

where $D(\mathbf{U}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\alpha}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{V}^{(t)}, \boldsymbol{\Omega}_1^{(t)}, \boldsymbol{\Omega}_2^{(t)})$ is calculated at each iteration and the tail-area of a χ^2 distribution which can be obtained via standard statistical software. The hypothesized model is rejected if $\hat{P}_B(\mathbf{X}, \mathbf{Z})$ is not close to 0.5.

Appendix 6.4: WinBUGS Code

```
model {
for (g in 1:G) {
#second level
for (j in 1:P) { vg[g,j]~dnorm(u2[g,j],psi2[j]) }

u2[g,1]<-1.0*xi2[g,1]
u2[g,2]<-lb[1]*xi2[g,1]
u2[g,3]<-lb[2]*xi2[g,1]

u2[g,4]<-1.0*xi2[g,2]
u2[g,5]<-lb[3]*xi2[g,2]
u2[g,6]<-lb[4]*xi2[g,2]

u2[g,7]<-1.0*xi2[g,3]
u2[g,8]<-lb[5]*xi2[g,3]
```

```
u2[g,9]<-lb[6]*xi2[g,3]

xi2[g,1:3]~dmnorm(ux2[1:3],phip[1:3,1:3])

#first model
for (i in 1:N[g]) {

for (j in 1:6) {
w[kk[g]+i,j]~dcat(p[kk[g]+i,j,1:C])
p[kk[g]+i,j,1]<-Q[kk[g]+i,j,1]
for (t in 2:C-1) { p[kk[g]+i,j,t]<-Q[kk[g]+i,j,t]-Q[kk[g]+i,j,t-1] }
p[kk[g]+i,j,C]<-1-Q[kk[g]+i,j,C-1]
for (t in 1:C-1)
{ Q[kk[g]+i,j,t]<-phi((alph[j,t]-u1[kk[g]+i,j])*sqrt(psi11[j])) }
}
#Note: the coding for the ordered categorical variable is 1,2,...,C

for (j in 7:9) { w[kk[g]+i,j]~dnorm(u1[kk[g]+i,j],psi1[j]) }

u1[kk[g]+i,1]<-vg[g,1]+1.0*eta1[g,i]
u1[kk[g]+i,2]<-vg[g,2]+lw[1]*eta1[g,i]
u1[kk[g]+i,3]<-vg[g,3]+lw[2]*eta1[g,i]

u1[kk[g]+i,4]<-vg[g,4]+1.0*xi1[g,i,1]
u1[kk[g]+i,5]<-vg[g,5]+lw[3]*xi1[g,i,1]
u1[kk[g]+i,6]<-vg[g,6]+lw[4]*xi1[g,i,1]

u1[kk[g]+i,7]<-vg[g,7]+1.0*xi1[g,i,2]
u1[kk[g]+i,8]<-vg[g,8]+lw[5]*xi1[g,i,2]
u1[kk[g]+i,9]<-vg[g,9]+lw[6]*xi1[g,i,2]

#Structural Equation model
eta1[g,i]~dnorm(nu1[g,i], psd)
 nu1[g,i]<-gam[1]*xi1[g,i,1]+gam[2]*xi1[g,i,2]+gam[3]*xi1[g,i,1]*xi1[g,i,2]
          +gam[4]*xi1[g,i,1]*xi1[g,i,1]+gam[5]*xi1[g,i,2]*xi1[g,i,2]

xi1[g,i,1:2]~dmnorm(ux1[1:2],phi1[1:2,1:2])
} # end of i
```

```
} # end of g

for (i in 1:2) { ux1[i]<-0.0 }
for (i in 1:3) { ux2[i]<-0.0 }

    # priors on loadings and coefficients
    lb[1]~dnorm(lbp[1],psi2[2]) lb[2]~dnorm(lbp[2],psi2[3])
    lb[3]~dnorm(lbp[3],psi2[5]) lb[4]~dnorm(lbp[4],psi2[6])
    lb[5]~dnorm(lbp[5],psi2[8]) lb[6]~dnorm(lbp[6],psi2[9])

    lw[1]~dnorm(lwp[1],psi1[2]) lw[2]~dnorm(lwp[2],psi1[3])
    lw[3]~dnorm(lwp[3],psi1[5]) lw[4]~dnorm(lwp[4],psi1[6])
    lw[5]~dnorm(lwp[5],psi1[8]) lw[6]~dnorm(lwp[6],psi1[9])

for (i in 1:5) { gam[i]~dnorm(gamp[i], psd) }

# priors on thresholds
for (j in 1:6) {
alph[j,1]<-a
alph[j,2]~dnorm(0,0.01)I(alph[j,1],alph[j,3])
alph[j,3]~dnorm(0,0.01)I(alph[j,2],alph[j,4])
alph[j,4]<-b
}
# a, b are fixed to identify the ordered categorical variable

# priors on precisions

for (j in 1:9) {
psi1[j]~dgamma(10.0,8.0)
sgm1[j]<-1/psi1[j]
}

for (j in 1:6) { psi2[j]<-1.0/0.3 }

for (j in 7:9) {
psi2[j]~dgamma(10.0,8.0)
sgm2[j]<-1/psi2[j]
```

```
}

    psd~dgamma(10.0,8.0)
    sgd<-1/psd

    phi1[1:2,1:2]~dwish(R1[1:2,1:2],5)
    phx1[1:2,1:2]<-inverse(phi1[1:2,1:2])

    phi2[1:3,1:3]~dwish(R2[1:2,1:2],5)
    phx2[1:3,1:3]<-inverse(phi2[1:3,1:3])

} # end of model
```

Appendix 6.5: Conditional Distributions: Multisample SEMs

The conditional distributions $[\mathbf{\Omega}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}]$, $[\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}, \mathbf{\Omega}, \mathbf{X}, \mathbf{Z}]$, and $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \mathbf{\Omega}, \mathbf{X}, \mathbf{Z}]$ that are required in the implementation of the Gibbs sampler are presented in this appendix. Note that the results on the first two conditional distributions are natural extension of those given in Section 5.2, but they can be regarded as the special cases of those given in Section 6.2. Also note that we allow common parameters in $\boldsymbol{\theta}$ according to the constraints under the competing models.

Conditional distribution of $[\mathbf{\Omega}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}]$ can be obtained as below:

$$p[\mathbf{\Omega}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}] = \prod_{g=1}^{G}\prod_{i=1}^{N_g} p(\boldsymbol{\omega}_i^{(g)}|\mathbf{v}_i^{(g)}, \boldsymbol{\theta}^{(g)}),$$

where

$$p(\boldsymbol{\omega}_i^{(g)}|\mathbf{v}_i^{(g)}, \boldsymbol{\theta}^{(g)}) \propto \exp\left\{-\frac{1}{2}\left[(\mathbf{v}_i^{(g)} - \boldsymbol{\mu}^{(g)} - \mathbf{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)})^T {\mathbf{\Psi}_\epsilon^{(g)}}^{-1}(\mathbf{v}_i^{(g)} - \boldsymbol{\mu}^{(g)} - \mathbf{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)}) \right.\right.$$
$$\left.\left. +\ (\boldsymbol{\eta}_i^{(g)} - \mathbf{\Lambda}_\omega^{(g)}\mathbf{G}(\boldsymbol{\omega}_i^{(g)}))^T {\mathbf{\Psi}_\delta^{(g)}}^{-1}(\boldsymbol{\eta}_i^{(g)} - \mathbf{\Lambda}_\omega^{(g)}\mathbf{G}(\boldsymbol{\omega}_i^{(g)})) + {\boldsymbol{\xi}_i^{(g)}}^T {\mathbf{\Phi}^{(g)}}^{-1}\boldsymbol{\xi}_i^{(g)}\right]\right\}. \qquad (6.A10)$$

Since the conditional distribution of (6.A10) is not standard, the Metropolis-Hastings (MH) algorithm can be used to draw random observations from this distribution.

Under the multisample situation, the notation in the conditional distribution $[\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}, \mathbf{\Omega}, \mathbf{X}, \mathbf{Z}]$ is very tedious. The derivation is similar to the two-level case as given in Appendix 6.1, Equations (6.A3) and (6.A4). As $(\boldsymbol{\alpha}^{(g)}, \mathbf{Y}^{(g)})$ is independent with $(\boldsymbol{\alpha}^{(h)}, \mathbf{Y}^{(h)})$ for $g \neq h$, and $\boldsymbol{\Psi}_\delta^{(g)}$ is diagonal,

$$p(\boldsymbol{\alpha}, \mathbf{Y}|\cdot) = \prod_{g=1}^{G} p(\boldsymbol{\alpha}^{(g)}, \mathbf{Y}^{(g)}|\cdot) = \prod_{g=1}^{G}\prod_{k=1}^{s} p(\boldsymbol{\alpha}_k^{(g)}, \mathbf{Y}_k^{(g)}|\cdot), \tag{6.A11}$$

where $\mathbf{Y}_k^{(g)} = (y_{1k}^{(g)}, \cdots, y_{N_g k}^{(g)})^T$. Let $\boldsymbol{\Psi}_{\epsilon y}^{(g)}$, $\boldsymbol{\Lambda}_y^{(g)}$, and $\boldsymbol{\mu}_y^{(g)}$ be the submatrices and subvector of $\boldsymbol{\Psi}_\epsilon^{(g)}$, $\boldsymbol{\Lambda}^{(g)}$, and $\boldsymbol{\mu}^{(g)}$ corresponding to the ordered categorical variables in $\mathbf{Y}$; let $\psi_{\epsilon yk}^{(g)}$ be the kth diagonal element of $\boldsymbol{\Psi}_{\epsilon y}^{(g)}$, $\mu_{yk}^{(g)}$ be the kth element of $\boldsymbol{\mu}_y^{(g)}$, and $\boldsymbol{\Lambda}_{yk}^{(g)^T}$ be the kth row of $\boldsymbol{\Lambda}_y^{(g)}$, and $I_A(y)$ be an indicator function with value 1 if y in A and zero otherwise, $p(\boldsymbol{\alpha}, \mathbf{Y}|\cdot)$ can be obtained from (6.A11) and

$$p(\boldsymbol{\alpha}_k^{(g)}, \mathbf{Y}_k^{(g)}|\cdot) \propto \prod_{i=1}^{N_g} \phi\{\psi_{\epsilon yk}^{(g)-1/2}(y_{ik}^{(g)} - \mu_{yk}^{(g)} - \boldsymbol{\Lambda}_{yk}{}^{(g)T}\boldsymbol{\omega}_i^{(g)})\} I_{[\alpha_{k,z_{ik}}, \alpha_{k,z_{ik}+1}]}(y_{ik}^{(g)}), \tag{6.A12}$$

where ϕ is the probability density function of $N[0,1]$. Note that in (6.A12) the superscript '(g)' in the threshold is suppressed to simplify notation.

Under the prior distributions of components in $\boldsymbol{\theta}$ as given in Section 6.3, the conditional distribution $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z}]$ is presented. Note that as $\mathbf{Y}$ is given, the model is defined with continuous data; hence, the conditional distribution is independent of $\boldsymbol{\alpha}$ and $\mathbf{Z}$.

The conditional distribution of some components in $\boldsymbol{\theta}^{(g)}$, $g = 1, \cdots, G$, under the situation without any parameter constraints are given as follows. Let $\boldsymbol{\Lambda}_k^{(g)^T}$ be the kth row of $\boldsymbol{\Lambda}^{(g)}$, and $\psi_{\epsilon k}^{(g)}$ be the kth diagonal element of $\boldsymbol{\Psi}_\epsilon^{(g)}$, $\mathbf{V}_k^{*(g)T}$ be the kth row of $\mathbf{V}^{*(g)} = (\mathbf{v}_1^{(g)} - \boldsymbol{\mu}^{(g)}, \cdots, \mathbf{v}_{N_g}^{(g)} - \boldsymbol{\mu}^{(g)})$, and $\boldsymbol{\Omega}_2^{(g)} = (\boldsymbol{\xi}_1^{(g)}, \cdots, \boldsymbol{\xi}_{N_g}^{(g)})$. It can be shown that:

$$\begin{aligned}
&[\boldsymbol{\mu}^{(g)}|\boldsymbol{\Lambda}^{(g)}, \boldsymbol{\Psi}_\epsilon^{(g)}, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}] \stackrel{D}{=} N(\mathbf{a}_\mu^{(g)}, \mathbf{A}_\mu^{(g)}), \\
&[\boldsymbol{\Lambda}_k^{(g)}|\boldsymbol{\Psi}_\epsilon^{(g)}, \boldsymbol{\mu}^{(g)}, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}] \stackrel{D}{=} N(\mathbf{a}_k^{(g)}, \mathbf{A}_k^{(g)}), \\
&[\psi_{\epsilon k}^{(g)-1}|\boldsymbol{\Lambda}_k^{(g)}, \boldsymbol{\mu}^{(g)}, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}] \stackrel{D}{=} Gamma(N_g/2 + \alpha_{0\epsilon k}^{(g)}, \beta_{\epsilon k}^{(g)}), \\
&[\boldsymbol{\Phi}^{(g)}|\boldsymbol{\Omega}_2^{(g)}] \stackrel{D}{=} IW_{q_2}[(\boldsymbol{\Omega}_2^{(g)}\boldsymbol{\Omega}_2^{(g)T} + \mathbf{R}_0^{(g)-1}), N_g + \rho_0^{(g)}],
\end{aligned} \tag{6.A13}$$

in which

$$\mathbf{a}_\mu^{(g)} = \mathbf{A}_\mu^{(g)}\big[\mathbf{\Sigma}_0^{(g)-1}\boldsymbol{\mu}_0^{(g)} + N_g\mathbf{\Psi}_\epsilon^{(g)-1}(\bar{\mathbf{v}}^{(g)} - \mathbf{\Lambda}^{(g)}\bar{\boldsymbol{\omega}}^{(g)})\big],\ \mathbf{A}_\mu^{(g)} = \big(\mathbf{\Sigma}_0^{(g)-1} + N_g\mathbf{\Psi}_\epsilon^{(g)-1}\big)^{-1},$$

$$\mathbf{a}_k^{(g)} = \mathbf{A}_k^{(g)}\big[\mathbf{H}_{0yk}^{(g)-1}\mathbf{\Lambda}_{0k}^{(g)} + \psi_{\epsilon k}^{(g)-1}\mathbf{\Omega}^{(g)}\mathbf{V}_k^{*(g)}\big],\ \mathbf{A}_k^{(g)} = \big[\psi_{\epsilon k}^{(g)-1}\mathbf{\Omega}^{(g)}\mathbf{\Omega}^{(g)T} + \mathbf{H}_{0yk}^{(g)-1}\big]^{-1}$$

$$\beta_{\epsilon k}^{(g)} = \beta_{0\epsilon k}^{(g)} + \big[\mathbf{\Lambda}_k^{(g)T}\mathbf{\Omega}^{(g)}\mathbf{\Omega}^{(g)T}\mathbf{\Lambda}_k^{(g)} - 2\mathbf{\Lambda}_k^{(g)T}\mathbf{\Omega}^{(g)}\mathbf{V}_k^{*(g)} + \mathbf{V}_k^{*(g)T}\mathbf{V}_k^{*(g)}\big]/2,$$

with $\bar{\mathbf{v}}^{(g)} = \sum_{i=1}^{N_g}\mathbf{v}_i^{(g)}/N_g$ and $\bar{\boldsymbol{\omega}}^{(g)} = \sum_{i=1}^{N_g}\boldsymbol{\omega}_i^{(g)}/N_g$ are the means of $\mathbf{v}_i^{(g)}$ and $\boldsymbol{\omega}_i^{(g)}$ within the gth group.

As we mentioned, slight modifications are required to handle models with parameter constraints, see Section 6.3. Under the constraints $\mathbf{\Lambda}_k^{(1)} = \cdots = \mathbf{\Lambda}_k^{(G)} = \mathbf{\Lambda}_k$; the conjugate prior distribution of $\mathbf{\Lambda}_k$ is $N[\mathbf{\Lambda}_{0k}, \mathbf{H}_{0yk}]$, and the conditional distribution is

$$[\mathbf{\Lambda}_k|\psi_{\epsilon k}^{(1)}, \cdots, \psi_{\epsilon k}^{(G)}, \boldsymbol{\mu}^{(1)}, \cdots, \boldsymbol{\mu}^{(G)}, \mathbf{Y}, \mathbf{\Omega}, \mathbf{X}] \stackrel{D}{=} N[\mathbf{a}_k, \mathbf{A}_k], \tag{6.A14}$$

where $\mathbf{a}_k = \mathbf{A}_k(\mathbf{H}_{0yk}^{-1}\mathbf{\Lambda}_{0k} + \sum_{g=1}^{G} \psi_{\epsilon k}^{(g)-1}\,\mathbf{\Omega}^{(g)}\mathbf{V}_k^{*(g)})$, and $\mathbf{A}_k = (\sum_{g=1}^{G}\psi_{\epsilon k}^{(g)-1}\mathbf{\Omega}^{(g)}\mathbf{\Omega}^{(g)T} + \mathbf{H}_{0yk}^{-1})^{-1}$. Under the constraints $\psi_{\epsilon k}^{(1)} = \cdots = \psi_{\epsilon k}^{(G)} = \psi_{\epsilon k}$, the conjugate prior distribution of $\psi_{\epsilon k}^{-1}$ is $Gamma(\alpha_{0\epsilon k}, \beta_{0\epsilon k})$, and the conditional distribution is

$$[\psi_{\epsilon k}^{-1}|\mathbf{\Lambda}_k^{(1)}, \cdots, \mathbf{\Lambda}_k^{(G)}, \boldsymbol{\mu}^{(1)}, \cdots, \boldsymbol{\mu}^{(G)}, \mathbf{Y}, \mathbf{\Omega}, \mathbf{X}] \stackrel{D}{=} Gamma(N^*/2 + \alpha_{0\epsilon k}, \beta_{\epsilon k}), \tag{6.A15}$$

where $N^* = N_1 + \cdots + N_G$,

$$\beta_{\epsilon k} = \beta_{0\epsilon k} + \sum_{g=1}^{G}\big[\mathbf{\Lambda}_k^{(g)T}(\mathbf{\Omega}^{(g)}\mathbf{\Omega}^{(g)T})\mathbf{\Lambda}_k^{(g)} - 2\mathbf{\Lambda}_k^{(g)T}\mathbf{\Omega}^{(g)}\mathbf{V}_k^{*(g)} + \mathbf{V}_k^{*(g)T}\mathbf{V}_k^{*(g)}\big]/2.$$

Under the constraints $\mathbf{\Phi}^{(1)} = \cdots = \mathbf{\Phi}^{(G)} = \mathbf{\Phi}$, the conjugate prior distribution of $\mathbf{\Phi}^{-1}$ is $W_{q_2}[\mathbf{R}_0, \rho_0]$, and the conditional distribution is

$$[\mathbf{\Phi}|\mathbf{\Omega}_2^{(1)}, \cdots, \mathbf{\Omega}_2^{(G)}] \stackrel{D}{=}\ IW_{q_2}[(\sum_{g=1}^{G}\mathbf{\Omega}_2^{(g)}\mathbf{\Omega}_2^{(g)T} + \mathbf{R}_0^{-1}),\ N^* + \rho_0]. \tag{6.A16}$$

The conditional distributions of $\mathbf{\Lambda}_{\omega k}^{(g)}$ and $\psi_{\delta k}^{(g)}$ are similar, and hence not presented.

As the conditional distributions involved in (6.A13) or (6.A14)-(6.A16) are standard distributions, drawing observations from them is straightforward. Simulating observations from the conditional distributions that are given in (6.A12) involves the univariate truncated normal distribution, and this is done by the inverse distribution method proposed by Robert (1995). A Metropolis-Hastings (MH) algorithm is used to simulate observations from the more complex conditional distribution (6.A10).

References

Cowles, M. K. (1996) Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, **6**, 101-111.

Gelman, A., Meng, X. L. and Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**, 733-760.

Gelman, A., Roberts, G. O. and Gilks, W. R. (1996) Efficient Metropolis jumping rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 5*, pp. 599-607. Oxford: Oxford University Press.

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

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

Kass, R. E. and Raftery, A. E. (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.

Lee, S. Y. (2007) *Structural Equation Modeling: A Bayesian Approach.* UK: John Wiley & Sons, Ltd.

Lee, S. Y. and Song, X. Y. (2005) Maximum likelihood analysis of a two-level nonlinear structural equation model with fixed covariates. *Journal of Educational and Behavioral Statistics*, **30**, 1-26.

Lee, S. Y. and Zhu, H. T. (2000) Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, **53**, 209-232.

Lee, S. Y., Poon, W. Y. and Bentler, P. M. (1989) Simultaneous analysis of multivariate polytomous variates in several groups. *Psychometrika*, **54**, 63-73.

Lee, S. Y., Song, X. Y. and Tang, N. S. (2007) Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling - A Multidisciplinary Journal*, **14**, 404-434.

Lee, S. Y., Song, X. Y., Skevington, S. and Hao, Y. T. (2005) Application of structural equation models to quality of life. *Structural Equation Modeling - A Multidisciplinary Journal*, **12**, 435-453.

McDonald, R. P. and Goldstein, H. (1989) Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, **42**, 215-232.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, **21**, 1087-1092.

Morisky, D. E., Tiglao, T. V., Sneed, C. D., Tempongko, S. B., Baltazar, J. C., Detels, R. and Stein, J.A. (1998) The effects of establishment practices, knowledge and attitudes on condom use among Filipina sex workers. *AIDS Care*, **10**, 213-220.

Power, M., Bullinger, M., Harper, A. and WHOQOL Group. (1999) The World Health Organization WHOQOL-100: Tests of the universality of quality of life in 15 different cultural groups worldwide. *Health Psychology*, **18**, 495-505.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004) Generalized multilevel structural equation modeling. *Psychometrika*, **69**, 167-190.

Robert, C. P. (1995) Simulation of truncated normal variables. *Statistics and Computing*, **5**, 121-125.

Shi, J. Q. and 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.

Shi, J. Q. and Lee, S. Y. (2000) Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society, Series B*, **62**, 77-87.

Skrondal, A. and Rabe-Hesketh, S. (2004) *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Florida: Chapman and Hall/CRC.

Song, X. Y. and Lee, S. Y. (2001) Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical and Statistical Psychology*, **54** 237-263.

Song, X. Y. and Lee, S. Y. (2004) Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, **57**, 29-52.

Song, X. Y. and Lee, S. Y. (2007) Bayesian analysis of latent variable models with non-ignorable missing outcomes from exponential family. *Statistics in Medicine*, **26**, 681-693.

Spiegelhalter, D. J., Thomas, A., Best, N. G. and Lunn, D. (2003) *WinBUGS User Manual. Version 1.4*. Cambridge, England: MRC Biostatistics Unit.

Zhu, H. T. and Lee, S. Y. (1999) Statistical analysis of nonlinear factor analysis models. *British Journal of Mathematical and Statistical Psychology*, **52**, 225-242.

Table 6.1: Bayesian estimates of the structural parameters and thresholds under Prior (I) for M_1: AIDS Data.

	Within Group			Between Group	
	EST	SE		EST	SE
Str. Par			Str. Par		
$\lambda_{1,21}$	0.238	0.081	$\lambda_{2,21}$	1.248	0.218
$\lambda_{1,31}$	0.479	0.112	$\lambda_{2,31}$	0.839	0.189
$\lambda_{1,52}$	1.102	0.213	$\lambda_{2,52}$	0.205	0.218
$\lambda_{1,62}$	0.973	0.185	$\lambda_{2,62}$	0.434	0.221
$\lambda_{1,83}$	0.842	0.182	$\lambda_{2,83}$	0.159	0.209
$\lambda_{1,93}$	0.885	0.192	$\lambda_{2,93}$	0.094	0.164
γ_{11}	0.454	0.147	$\phi_{2,11}$	0.212	0.042
γ_{12}	-0.159	0.159	$\phi_{2,12}$	-0.032	0.032
γ_{13}	-0.227	0.382	$\phi_{2,13}$	0.008	0.037
$\phi_{1,11}$	0.216	0.035	$\phi_{2,22}$	0.236	0.054
$\phi_{1,12}$	-0.031	0.017	$\phi_{2,23}$	0.006	0.041
$\phi_{1,22}$	0.202	0.037	$\phi_{2,33}$	0.257	0.063
ψ_{11}	0.558	0.087	ψ_{27}	0.378	0.070
ψ_{12}	0.587	0.049	ψ_{28}	0.349	0.053
ψ_{13}	0.725	0.063	ψ_{29}	0.259	0.039
ψ_{14}	0.839	0.084			
ψ_{15}	0.691	0.085			
ψ_{16}	0.730	0.081			
ψ_{17}	0.723	0.056			
ψ_{18}	0.629	0.053			
ψ_{19}	0.821	0.062			
$\psi_{1\delta}$	0.460	0.080			
Thresholds					
α_{12}	-1.163	0.054	α_{13}	-0.751	0.045
α_{22}	-0.083	0.033	α_{23}	0.302	0.035
α_{32}	-0.985	0.045	α_{33}	-0.589	0.044
α_{42}	-0.406	0.035	α_{43}	0.241	0.029
α_{52}	-1.643	0.063	α_{53}	-0.734	0.027
α_{62}	-1.038	0.043	α_{63}	-0.118	0.025

This table and Tables 6.2 and 6.3 are extracted from Lee (2007).

Table 6.2: Bayesian estimates of the structural parameters and thresholds under Prior (II) for M_1: AIDS Data.

	Within Group			Between Group	
	EST	SE		EST	SE
Str. Par			Str. Par		
$\lambda_{1,21}$	0.239	0.080	$\lambda_{2,21}$	1.404	0.283
$\lambda_{1,31}$	0.495	0.119	$\lambda_{2,31}$	0.869	0.228
$\lambda_{1,52}$	1.210	0.284	$\lambda_{2,52}$	0.304	0.293
$\lambda_{1,62}$	1.083	0.250	$\lambda_{2,62}$	0.602	0.264
$\lambda_{1,83}$	0.988	0.215	$\lambda_{2,83}$	0.155	0.230
$\lambda_{1,93}$	0.918	0.197	$\lambda_{2,93}$	0.085	0.176
γ_{11}	0.474	0.157	$\phi_{2,11}$	0.196	0.043
γ_{12}	-0.232	0.165	$\phi_{2,12}$	-0.026	0.030
γ_{13}	-0.353	0.540	$\phi_{2,13}$	0.010	0.032
$\phi_{1,11}$	0.198	0.048	$\phi_{2,22}$	0.219	0.048
$\phi_{1,12}$	-0.022	0.015	$\phi_{2,23}$	0.008	0.037
$\phi_{1,22}$	0.181	0.035	$\phi_{2,33}$	0.251	0.060
ψ_{11}	0.562	0.100	ψ_{27}	0.376	0.068
ψ_{12}	0.587	0.049	ψ_{28}	0.350	0.055
ψ_{13}	0.715	0.066	ψ_{29}	0.256	0.039
ψ_{14}	0.849	0.091			
ψ_{15}	0.702	0.093			
ψ_{16}	0.697	0.079			
ψ_{17}	0.738	0.055			
ψ_{18}	0.601	0.056			
ψ_{19}	0.828	0.064			
$\psi_{1\delta}$	0.460	0.077			
Thresholds					
α_{12}	-1.170	0.060	α_{13}	-0.755	0.053
α_{22}	-0.084	0.029	α_{23}	0.303	0.033
α_{32}	-0.986	0.043	α_{33}	-0.589	0.042
α_{42}	-0.406	0.035	α_{43}	0.242	0.029
α_{52}	-1.656	0.066	α_{53}	-0.738	0.028
α_{62}	-1.032	0.044	α_{63}	-0.119	0.026

Table 6.3: Bayesian estimates of the structural parameters and thresholds under Prior (I) for M_4: AIDS Data.

	Within Group			Between Group	
	EST	SE		EST	SE
Str. Par			Str. Par		
$\lambda_{1,21}$	0.203	0.070	$\lambda_{2,21}$	1.261	0.233
$\lambda_{1,31}$	0.450	0.100	$\lambda_{2,31}$	0.842	0.193
$\lambda_{1,52}$	0.992	0.205	$\lambda_{2,52}$	0.189	0.227
$\lambda_{1,62}$	0.868	0.180	$\lambda_{2,62}$	0.461	0.209
$\lambda_{1,83}$	0.936	0.172	$\lambda_{2,83}$	0.157	0.230
$\lambda_{1,93}$	0.880	0.194	$\lambda_{2,93}$	0.074	0.167
γ_{11}	0.489	0.147	$\phi_{2,11}$	0.211	0.040
γ_{12}	-0.026	0.217	$\phi_{2,12}$	-0.029	0.033
γ_{13}	-0.212	0.265	$\phi_{2,13}$	0.010	0.035
γ_{14}	0.383	0.442	$\phi_{2,22}$	0.223	0.053
γ_{15}	-0.147	0.188	$\phi_{2,23}$	0.013	0.038
$\phi_{1,11}$	0.245	0.042	$\phi_{2,33}$	0.243	0.059
$\phi_{1,12}$	-0.029	0.020	ψ_{27}	0.377	0.068
$\phi_{1,22}$	0.186	0.031	ψ_{28}	0.351	0.055
ψ_{11}	0.546	0.093	ψ_{29}	0.258	0.040
ψ_{12}	0.591	0.047			
ψ_{13}	0.724	0.063			
ψ_{14}	0.826	0.081			
ψ_{15}	0.716	0.092			
ψ_{16}	0.731	0.077			
ψ_{17}	0.733	0.049			
ψ_{18}	0.610	0.048			
ψ_{19}	0.833	0.064			
$\psi_{1\delta}$	0.478	0.072			
Thresholds					
α_{12}	-1.163	0.058	α_{13}	-0.753	0.048
α_{22}	-0.088	0.032	α_{23}	0.299	0.036
α_{32}	-0.980	0.046	α_{33}	-0.583	0.047
α_{42}	-0.407	0.034	α_{43}	0.244	0.028
α_{52}	-1.650	0.063	α_{53}	-0.734	0.027
α_{62}	-1.034	0.043	α_{63}	-0.118	0.026

Table 6.4: Bayesian estimates of unknown parameters in the 2-group SEM with no constraints.

	Group			Group			Group	
	g=1	g=2		g=1	g=2		g=1	g=2
μ_1	0.021	-0.519	$\lambda_{2,1}$	0.859	0.804	γ_1	0.847	0.539
μ_2	0.001	0.059	$\lambda_{4,2}$	0.952	0.754	γ_2	0.334	0.139
μ_3	0.002	-0.240	$\lambda_{5,2}$	1.112	1.016	γ_3	0.167	0.026
μ_4	0.009	-0.300	$\lambda_{6,2}$	1.212	0.976	γ_4	-0.068	0.241
μ_5	-0.004	-0.188	$\lambda_{7,2}$	0.820	0.805	ψ_1	0.400	0.248
μ_6	0.008	-0.382	$\lambda_{8,2}$	1.333	1.123	ψ_2	0.422	0.268
μ_7	-0.002	-0.030	$\lambda_{9,2}$	1.203	0.961	ψ_3	0.616	0.584
μ_8	0.008	-0.070	$\lambda_{11,3}$	0.799	0.827	ψ_4	0.628	0.445
μ_9	0.003	0.108	$\lambda_{12,3}$	0.726	0.987	ψ_5	0.462	0.214
μ_{10}	0.004	-0.358	$\lambda_{13,3}$	0.755	0.669	ψ_6	0.401	0.184
μ_{11}	0.003	-0.286	$\lambda_{14,3}$	1.011	0.762	ψ_7	0.709	0.253
μ_{12}	0.001	-0.087	$\lambda_{15,3}$	0.874	0.719	ψ_8	0.271	0.202
μ_{13}	0.004	-0.373	$\lambda_{17,4}$	0.273	0.627	ψ_9	0.393	0.191
μ_{14}	0.003	0.031	$\lambda_{18,4}$	0.954	0.961	ψ_{10}	0.471	0.288
μ_{15}	0.002	0.079	$\lambda_{20,5}$	0.804	1.108	ψ_{11}	0.654	0.262
μ_{16}	0.012	-0.404	$\lambda_{21,5}$	0.772	0.853	ψ_{12}	0.707	0.428
μ_{17}	0.000	0.037	$\lambda_{22,5}$	0.755	0.815	ψ_{13}	0.698	0.348
μ_{18}	0.010	-0.596	$\lambda_{23,5}$	0.723	0.672	ψ_{14}	0.453	0.269
μ_{19}	0.005	-0.183	$\lambda_{24,5}$	0.984	0.647	ψ_{15}	0.575	1.137
μ_{20}	0.004	-0.543	$\lambda_{25,5}$	0.770	0.714	ψ_{16}	0.462	0.267
μ_{21}	0.003	-0.571	$\lambda_{26,5}$	0.842	0.761	ψ_{17}	0.962	0.297
μ_{22}	0.002	-0.966	ϕ_{11}	0.450	0.301	ψ_{18}	0.522	0.301
μ_{23}	0.001	-0.220	ϕ_{12}	0.337	0.279	ψ_{19}	0.530	0.559
μ_{24}	0.017	-1.151	ϕ_{13}	0.211	0.162	ψ_{20}	0.679	0.565
μ_{25}	-0.001	-0.837	ϕ_{14}	0.299	0.207	ψ_{21}	0.708	0.392
μ_{26}	0.007	-0.982	ϕ_{22}	0.579	0.537	ψ_{22}	0.714	0.386
			ϕ_{23}	0.390	0.251	ψ_{23}	0.736	0.493
			ϕ_{24}	0.393	0.290	ψ_{24}	0.577	0.451
			ϕ_{33}	0.599	0.301	ψ_{25}	0.719	0.482
			ϕ_{34}	0.386	0.210	ψ_{26}	0.670	0.408
			ϕ_{44}	0.535	0.386			
			ψ_δ	0.246	0.234			