Chapter 6 Structural Equation Models for Analyzing Hierarchical and Multisample Data

In many substantive researches, one may encounter the following types of data:

1. mixture data — 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.
2. hierarchical data — come from a number of different groups (clusters) with a known hierarchical structure. Examples include:
   - patients from within random samples of clinics or hospitals;
   - individuals from within random samples of families;
   - students from within random samples of schools.
3. multisample data — come from a number of distinct groups (populations), where the number of groups is known, and the group membership of each observation can be specified exactly.

Differences between mixture, multisapmle, and hierarchical data:

1. K is usually small in mixture and multisample data, but is large for hierarchical data. The number of observations within each group is large for mixture and multisample data, but is relatively small for hierarchical data.
2. The group membership of each observation is unknown for mixture data, but is specified for hierarchical and multisample data.
3. The observations in each group are independent for mixture and multisample data, but dependent for hierarchical data because individuals within a group share certain common influential factors.

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:

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

$y_{ij} = \alpha + \beta x_{ij} + u_j + \varepsilon_{ij}$. What's j ?

Why not build G separated models for every group?

where

- $\mathbf{\Lambda}_{1g}$ — $p \times q_1$ matrix of factor loadings,
- $\boldsymbol{\omega}_{1gi}$ — $q_1 \times 1$ random vector of latent factors,
- $\boldsymbol{\epsilon}_{1gi}$ — $p \times 1$ 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 dependent due to the existence of $\mathbf{v}_g$. Hence, in the two-level SEM, the usual assumption on the independence of 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:

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

where

- $\boldsymbol{\mu}$ — vector of intercepts,
- $\boldsymbol{\Lambda}_2$ — $p \times q_2$ matrix of factor loadings,
- $\boldsymbol{\omega}_{2g}$ — $q_2 \times 1$ vector of latent variables,
- $\boldsymbol{\epsilon}_{2g}$ — $p \times 1$ 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, $\boldsymbol{\epsilon}_{2g}$ is assumed to be independent of $\boldsymbol{\epsilon}_{1gi}$.

It follows from equations (1) and (2) that

group level error

individual level error.

$$\mathbf{u}_{gi} = \underbrace{\boldsymbol{\mu} + \boldsymbol{\Lambda}_2 \boldsymbol{\omega}_{2g}}_{\text{common effect}} + \boldsymbol{\epsilon}_{2g} + \underbrace{\boldsymbol{\Lambda}_{1g} \boldsymbol{\omega}_{1gi}}_{\text{group specific effect.}} + \boldsymbol{\epsilon}_{1gi}. \tag{3}$$

Let $\boldsymbol{\omega}_{1gi} = (\boldsymbol{\eta}_{1gi}^T, \boldsymbol{\xi}_{1gi}^T)^T$, $\boldsymbol{\omega}_{2g} = (\boldsymbol{\eta}_{2g}^T, \boldsymbol{\xi}_{2g}^T)^T$. Consider

$$\boldsymbol{\eta}_{1gi} = \boldsymbol{\Pi}_{1g}\boldsymbol{\eta}_{1gi} + \boldsymbol{\Gamma}_{1g}\mathbf{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}, \quad \text{and} \tag{4}$$

$$\boldsymbol{\eta}_{2g} = \boldsymbol{\Pi}_2\boldsymbol{\eta}_{2g} + \boldsymbol{\Gamma}_2\mathbf{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g}, \tag{5}$$

where $\mathbf{F}_1, \mathbf{F}_2, \boldsymbol{\Pi}_{1g}, \boldsymbol{\Pi}_2, \boldsymbol{\Gamma}_{1g}, \boldsymbol{\Gamma}_2$ are defined as before, and $\boldsymbol{\xi}_{1gi} \sim N[\mathbf{0}, \boldsymbol{\Phi}_{1g}]$, $\boldsymbol{\xi}_{2g} \sim N[\mathbf{0}, \boldsymbol{\Phi}_2]$, $\boldsymbol{\delta}_{1gi} \sim N[\mathbf{0}, \boldsymbol{\Psi}_{1g\delta}]$, $\boldsymbol{\delta}_{2g} \sim N[\mathbf{0}, \boldsymbol{\Psi}_{2\delta}]$.

Remarks:

1. Due to the nonlinearity induced by $\mathbf{F}_1$ and $\mathbf{F}_2$, the underlying distribution of $\mathbf{u}_{gi}$ is not normal.
2. $\boldsymbol{\eta}_{1gi}$ 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.
3. The method of fixing appropriate elements in $\boldsymbol{\Lambda}_{1g}, \boldsymbol{\Pi}_{1g}, \boldsymbol{\Gamma}_{1g}$, $\boldsymbol{\Lambda}_2, \boldsymbol{\Pi}_2$, and $\boldsymbol{\Gamma}_2$ at preassigned values can be used to identify the two-level SEM.

To accommodate mixed ordered categorical and continuous variables, 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 an observable continuous random vector, and $\mathbf{y}_{gi} = (y_{gi1}, \cdots, y_{gis})^T$ is a unobservable continuous random vector, and corresponds to the observable ordered categorical vector $\mathbf{z} = (z_1, \cdots, z_s)^T$ through a threshould specification.

Note that

1. The model is general, including three major components: (1) a two-level model for hierarchically structured data, (2) discrete data, (3) nonlinear structural equations in the within-group and between-group models.

- When G is not large, we usually consider a simple between-group model.
- Most two-level SEMs in the literature assume that the within-group parameters are invariant over groups. $\Lambda_g \equiv \Lambda \quad g=1,\ldots,G$?

Let

- θ — vector of unknown structural parameters
- α — vector of unknown thresholds
- $\mathbf{X}_g = (\mathbf{x}_{g1}, \cdots, \mathbf{x}_{gN_g})$, $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_G)$ — observed continuous data
- $\mathbf{Z}_g = (\mathbf{z}_{g1}, \cdots, \mathbf{z}_{gN_g})$, $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_G)$ — observed ordered categorical data,
- $\mathbf{Y}_g = (\mathbf{y}_{g1}, \cdots, \mathbf{y}_{gN_g})$, $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_G)$ — latent continuous data associated with $\mathbf{Z}_g$ and $\mathbf{Z}$,
- $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_G)$ — matrix of between-group latent variables.
- $\boldsymbol{\Omega}_{1g} = (\boldsymbol{\omega}_{1g1}, \cdots, \boldsymbol{\omega}_{1gN_g}), \boldsymbol{\Omega}_1 = (\boldsymbol{\Omega}_{11}, \cdots, \boldsymbol{\Omega}_{1G}), \boldsymbol{\Omega}_2 = (\boldsymbol{\omega}_{21}, \cdots, \boldsymbol{\omega}_{2G})$ — matrices of latent variables at the within-group and between-group levels.

In the posterior analysis, we consider the joint posterior distribution $[\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2 | \mathbf{X}, \mathbf{Z}]$. With the Gibbs sampler, we iteratively sample from the following conditional distributions:

$$
\begin{aligned}
&[\mathbf{V}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}], \quad [\boldsymbol{\Omega}_1|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}],\\
&[\boldsymbol{\Omega}_2|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \mathbf{X}, \mathbf{Z}], \quad [\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}],\\
&[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}].
\end{aligned}
$$

For the two-level model, $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \mathbf{X}, \mathbf{Z}]$ is further decomposed into components involving parameters in the between-group and within-group models. These components are different under various special cases of the model. Some typical examples are given below.

(A) Models with different within-group parameters across groups: in this case, $\boldsymbol{\theta}_{1g} = \{\boldsymbol{\Lambda}_{1g}, \boldsymbol{\Psi}_{1g}, \boldsymbol{\Pi}_{1g}, \boldsymbol{\Gamma}_{1g}, \boldsymbol{\Phi}_{1g}, \boldsymbol{\Psi}_{1g\delta}\}$ and $\boldsymbol{\alpha}_g$ are different from $\boldsymbol{\alpha}_h$, 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.

the model of each level should have sufficient data size, i.e G for between group effect, N_g for within group effect.

(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: in this situation, $\boldsymbol{\theta}_{11} = \cdots = \boldsymbol{\theta}_{1G}$, and $\boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_G$.

The conditional distributions of the components in $[\theta|\alpha, \mathbf{Y}, \mathbf{V}, \Omega_1, \Omega_2, \mathbf{X}, \mathbf{Z}]$ as well as other conditional distributions required by the Gibbs sampler are briefly discussed in Appendix 6.1. Some technical details on the implementation of the MH algorithm are given in Appendix 6.2.

The Bayes factor is used for comparing competing M_0 and M_1:

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

The path sampling is again used to compute B_{10}. Here,

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

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

$$\mathbf{u}_{gi} = \mathbf{v}_g + \boldsymbol{\Lambda}_1 \boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}, \quad (6)$$

$$\boldsymbol{\eta}_{1gi} = \boldsymbol{\Pi}_1 \boldsymbol{\eta}_{1gi} + \boldsymbol{\Gamma}_1 \mathbf{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}. \quad (7)$$

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

Let

$$M_1: \ \mathbf{v}_g = \boldsymbol{\mu} + \boldsymbol{\Lambda}_2^1 \boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}, \tag{8}$$

where $\boldsymbol{\omega}_{2g}$ is distributed as $N[\mathbf{0}, \boldsymbol{\Phi}_2]$. 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:

$$\begin{aligned} M_2: \quad \mathbf{v}_g &= \boldsymbol{\mu} + \boldsymbol{\Lambda}_2^2 \boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}, & (9) \\ \boldsymbol{\eta}_{2g} &= \boldsymbol{\Pi}_2^2 \boldsymbol{\eta}_{2g} + \boldsymbol{\Gamma}_2^2 \mathbf{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g}. & (10) \end{aligned}$$

The between-group model in M_2 is a nonlinear SEM. 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. Then,

$$\log B_{12} = \log \frac{p(\mathbf{X}, \mathbf{Z}|M_1)/p(\mathbf{X}, \mathbf{Z}|M_a)}{p(\mathbf{X}, \mathbf{Z}|M_2)/p(\mathbf{X}, \mathbf{Z}|M_a)} = \log B_{1a} - \log B_{2a}. \tag{11}$$

One auxiliary model is M_a, in which the measurement and structural equations of the within-group model are:

$$\begin{aligned} \mathbf{u}_{gi} &= \mathbf{v}_g + \boldsymbol{\Lambda}_1 \boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}, \\ \boldsymbol{\eta}_{1gi} &= \boldsymbol{\Pi}_1 \boldsymbol{\eta}_{1gi} + \boldsymbol{\Gamma}_1 \mathbf{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}. \end{aligned}$$

and the between-group model is defined by

$$\mathbf{v}_g = \boldsymbol{\mu} + \boldsymbol{\epsilon}_{2g}.$$

The link model M_{t1a} is defined by

between group effect.

$$M_{t1a}: \ \mathbf{u}_{gi} = \boldsymbol{\mu} + t\boldsymbol{\Lambda}_2^1 \boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g} + \boldsymbol{\Lambda}_1 \boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi},$$

with the within-group structural equation given by

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

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

$$\begin{aligned} \boldsymbol{\eta}_{1gi} &= \boldsymbol{\Pi}_1\boldsymbol{\eta}_{1gi} + \boldsymbol{\Gamma}_1\mathbf{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}, \\ \boldsymbol{\eta}_{2g} &= \boldsymbol{\Pi}_2^2\boldsymbol{\eta}_{2g} + \boldsymbol{\Gamma}_2^2\mathbf{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g}. \end{aligned}$$

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 as follows:

$$\log B_{12} = \log B_{1a} - \log B_{2a}.$$

As an illustration, we use a small portion of the data 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).

Data and variables:

- 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.
- The sample size is 755 after deleting missing entries.
- There are 97 establishments. The numbers of individuals in establishments varied from 1 to 58; this gives an unbalanced data set.
- The first three, the next three, and the last three observed variables are used as indicators of latent factors:'worry about AIDS', 'attitude to the risk of getting AIDS', and 'aggressiveness'.
- Establishment policies have a strong influence on CSWs, so the observations within each establishment are correlated.

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. The non-overlapping structure of $\mathbf{\Lambda}_2$ gives clear interpretation of latent factors.

For the within-group model, we considered invariant within-group parameters such that

$$\mathbf{\Psi}_{1g} = \mathbf{\Psi}_1 = \mathrm{diag}(\psi_{11}, \cdots, \psi_{19}), \quad \mathbf{\Lambda}_{1g} = \mathbf{\Lambda}_1,$$

where $\mathbf{\Lambda}_1$ has the same common structure as $\mathbf{\Lambda}_2$.

To assess the interaction effect of the explanatory latent factors, the following structural equation is considered:

$$\eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{13}\xi_{1gi1}\xi_{1gi2} + \delta_{1gi}. \qquad (12)$$

To identity the model with respect to ordered categorical variables via the common method, α_{k1} and α_{k4}, $k = 1, \cdots, 6$ are fixed at $\alpha_{kj} = \Phi^{*^{-1}}(m_k)$, where m_k is the observed cumulative marginal proportion of the categories with $z_{gk} < j$. The Bayesian estimates of unknown parameters obtained under the conjugate prior inputs are reported in the following Table 6.1.

	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			

To illustrate the model comparison using Bayes factor, we compare this two-level nonlinear model with some non-nested and nested models.

Let M_1, M_2, and M_3 be non-nested models with the same measurement equation described above and different nonlinear structural equations:

$$\begin{aligned}
M_1: &\quad \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{13}\xi_{1gi1}\xi_{1gi2} + \delta_{1gi}, \\
M_2: &\quad \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{14}\xi^2_{1gi1} + \delta_{1gi}, \\
M_3: &\quad \eta_{1gi} = \gamma_{11}\xi_{1gi1} + \gamma_{12}\xi_{1gi2} + \gamma_{15}\xi^2_{1gi2} + \delta_{1gi}.
\end{aligned}$$

To compare M_1 and M_2, we use link model M_t as follows:

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

Clearly, when $t = 1$, $M_t = M_2$; when $t = 0$, $M_t = M_1$. The $\widehat{\log B_{21}}$, $\widehat{\log B_{23}}$ under prior inputs (I, II) are equal to (0.317,0.018) and (0.176, 0.131), respectively. Hence, 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 structural equations are given by:

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

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

The $\widehat{\log B_{40}}$ and $\widehat{\log B_{42}}$ 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, indicating that the selected model fits the data. The Bayesian estimates and their standard error estimates under M_4 and Prior inputs (I) are reported in Table 6.3 (not reported here).

Compared to hierarchical data, multisample data have different features:

(1) come from a comparatively smaller number of groups (populations);

(2) the number of observations within each group is usually large;

(3) observations within each group are assumed independent.

One of the main objectives in the analysis of multisample data is to investigate the similarities or differences among the models in the different groups. For example,

- investigate the behaviors of different groups of employees,
- compare different cultures,
- examine the drug effect in different treatment groups.

Hence, an important issue is to test hypotheses about the 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.

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:

$$\mathbf{v}_i^{(g)} = \boldsymbol{\mu}^{(g)} + \boldsymbol{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \tag{13}$$

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 $\boldsymbol{\epsilon}_i^{(g)} \sim N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon^{(g)}]$, where $\boldsymbol{\Psi}_\epsilon^{(g)}$ is diagonal.

Let $\omega_i^{(g)} = (\eta_i^{(g)T}, \xi_i^{(g)T})^T$. We consider the nonlinear structural equation:

$$\eta_i^{(g)} = \Pi^{(g)}\eta_i^{(g)} + \Gamma^{(g)}\mathbf{F}(\xi_i^{(g)}) + \delta_i^{(g)}. \tag{14}$$

It is assumed that just dimension

- the dimensions of $\xi_i^{(g)}$ and $\eta_i^{(g)}$ are independent of g; that is, they are the same for each group;
- $\xi_i^{(g)}$ and $\delta_i^{(g)}$ are independent;
- $\xi_i^{(g)} \sim N[\mathbf{0}, \Phi^{(g)}]$ and $\delta_i^{(g)} \sim N[\mathbf{0}, \Psi_\delta^{(g)}]$, where $\Psi_\delta^{(g)}$ is diagonal;
- $\mathbf{F}(\cdot)$ does not depend on g, but different groups can have different linear or nonlinear terms of $\xi_i^{(g)}$ by defining appropriate $\mathbf{F}(\cdot)$ and assigning zero values to appropriate elements in $\Gamma^{(g)}$.

Let $\Lambda_\omega^{(g)} = (\Pi^{(g)}, \Gamma^{(g)})$ and $\mathbf{G}(\omega_i^{(g)}) = (\eta_i^{(g)^T}, \mathbf{F}(\xi_i^{(g)})^T)^T$, (14) becomes

$$\eta_i^{(g)} = \Lambda_\omega^{(g)}\mathbf{G}(\omega_i^{(g)}) + \delta_i^{(g)}. \tag{15}$$

To handle ordered categorical outcomes, suppose that $\mathbf{v}_i^{(g)} = (\mathbf{x}_i^{(g)T}, \mathbf{y}_i^{(g)T})^T$, where

- $\mathbf{x}_i^{(g)}$ — $r \times 1$ subvector of observable continuous responses;
- $\mathbf{y}_i^{(g)}$ — $s \times 1$ subvector of unobservable continuous responses, the information of which is reflected by an observable ordered categorical vector $\mathbf{z}_i^{(g)}$ as follows:

$$z_m^{(g)} = a \quad \text{if} \quad \alpha_{m,a}^{(g)} \leq y_m^{(g)} < \alpha_{m,a+1}^{(g)}, \tag{16}$$

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 $z_m^{(g)}$. Here, b_m is independent of g. That is, for each ordered categorical variable, the number of thresholds is the same for each group.

same in each group

To tackle the identification problem related to the ordered categorical variables, we fix $\alpha_{m,1}^{(g)}$ and $\alpha_{m,b_m}^{(g)}$ at preassigned values. Another important issue is to impose conditions 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. Then, for any m, and $g \neq 1$, we impose the following restrictions:

$$\alpha_{m,k}^{(g)} = \alpha_{m,k}^{(1)}, \qquad k = 1, \cdots, b_m, \quad m = 1, \cdots, s. \tag{17}$$

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 not affected by the choice of the reference group. The compatibility of the groups is reflected by the differences of the parameter estimates.

Let

- $\boldsymbol{\theta} = \{\boldsymbol{\theta}^{(g)}, g = 1, \cdots, G\}$, where $\boldsymbol{\theta}^{(g)}$ is the vector of unknown parameters in the gth group;
- $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}^{(g)}, g = 1, \cdots, G\}$, where $\boldsymbol{\alpha}^{(g)}$ is the vector of unknown thresholds in the gth group;
- $\mathbf{X}^{(g)} = (\mathbf{x}_1^{(g)}, \cdots, \mathbf{x}_{N_g}^{(g)})$ and $\mathbf{X} = (\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(G)})$ — observed continuous data;
- $\mathbf{Z}^{(g)} = (\mathbf{z}_1^{(g)}, \cdots, \mathbf{z}_{N_g}^{(g)})$ and $\mathbf{Z} = (\mathbf{Z}^{(1)}, \cdots, \mathbf{Z}^{(G)})$ — observed ordered categorical data;
- $\mathbf{Y}^{(g)} = (\mathbf{y}_1^{(g)}, \cdots, \mathbf{y}_{N_g}^{(g)})$ and $\mathbf{Y} = (\mathbf{Y}^{(1)}, \cdots, \mathbf{Y}^{(G)})$ — latent continuous measurement associated with $\mathbf{Z}^{(g)}$ and $\mathbf{Z}$, respectively;
- $\boldsymbol{\Omega}^{(g)} = (\boldsymbol{\omega}_1^{(g)}, \cdots, \boldsymbol{\omega}_{N_g}^{(g)})$ and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}^{(1)}, \cdots, \boldsymbol{\Omega}^{(G)})$ — matrices of latent variables.

In multisample analysis, a certain type of parameter in $\boldsymbol{\theta}^{(g)}$ is often hypothesized to be invariant over the group models. For example, the following constraints are often imposed:

$$\boldsymbol{\Lambda}^{(1)} = \cdots = \boldsymbol{\Lambda}^{(G)}, \quad \boldsymbol{\Phi}^{(1)} = \cdots = \boldsymbol{\Phi}^{(G)}, \quad \boldsymbol{\Gamma}^{(1)} = \cdots = \boldsymbol{\Gamma}^{(G)}.$$

Testing these constraints can be formulated as the model comparison problem. Some examples of nonnested competing models (or hypotheses) are:

$$\begin{array}{lll} M_A: & \text{No constraints}, & \\ M_1: & \boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(G)}, & M_2: \ \boldsymbol{\Lambda}^{(1)} = \cdots = \boldsymbol{\Lambda}^{(G)}, \\ M_3: & \boldsymbol{\Lambda}_\omega^{(1)} = \cdots = \boldsymbol{\Lambda}_\omega^{(G)}, & M_4: \ \boldsymbol{\Phi}^{(1)} = \cdots = \boldsymbol{\Phi}^{(G)}, \\ M_5: & \boldsymbol{\Psi}_\epsilon^{(1)} = \cdots = \boldsymbol{\Psi}_\epsilon^{(G)}, & M_6: \ \boldsymbol{\Psi}_\delta^{(1)} = \cdots = \boldsymbol{\Psi}_\delta^{(G)}. \end{array} \tag{18}$$

In the posterior analysis, the components in $[\theta|\alpha, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z}]$ and the specification of prior distributions are slightly different under different M_k defined above.

1. Prior distributions for nonconstrained parameters in different groups are naturally assumed to be independent. So, in estimating the unconstrained parameters, we need to specify its own prior distribution, and the data in the corresponding group are used.
2. For constrained parameters, only one prior distribution for these constrained parameters is needed, and all the data in the groups should be combined in the estimation. Under this situation, we may not take a joint prior distribution for the factor loading matrix and the unique variance of the error measurement because this kind of joint prior distribution may cause problems under the constrained situation:

$$\boldsymbol{\Lambda}^{(1)} = \cdots = \boldsymbol{\Lambda}^{(G)} = \boldsymbol{\Lambda}, \quad \boldsymbol{\Psi}_\epsilon^{(1)} \neq \cdots \neq \boldsymbol{\Psi}_\epsilon^{(G)}.$$

Let $\boldsymbol{\Lambda}_{\omega k}^T$ and $\boldsymbol{\Lambda}_{\omega k}^{(g)^T}$ be the kth rows of $\boldsymbol{\Lambda}_{\omega}$ and $\boldsymbol{\Lambda}_{\omega}^{(g)}$, respectively. The prior distributions of $\boldsymbol{\Lambda}_{\omega}$ are given by:

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

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

Similarly, let $\psi_{\delta k}^{(g)}$ be the kth diagonal element of $\boldsymbol{\Psi}_{\delta}^{(g)}$. The prior distributions of $\psi_{\delta k}^{(g)-1}$ $\boldsymbol{\mu}^{(g)}$, and $\boldsymbol{\Phi}^{(g)}$ are

$$Gamma[\alpha_{0\delta k}^{(g)}, \beta_{0\delta k}^{(g)}], \quad \boldsymbol{\mu}^{(g)} \stackrel{D}{=} N[\boldsymbol{\mu}_0^{(g)}, \boldsymbol{\Sigma}_0^{(g)}], \quad \boldsymbol{\Phi}^{(g)-1} \stackrel{D}{=} W_{q_2}[\mathbf{R}_0^{(g)}, \rho_0^{(g)}].$$

Based on the specified priors, the conditional distribution $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z}]$ under various competing models can be obtained (see Appendix 6.5).

Let M_k and M_h be two nested or non-nested competing models (hypotheses), M_k and M_h can then be compared using the Bayes factor:

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

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

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

$H_0: \Lambda_1 = \cdots = \Lambda_G$

$H_1: \Gamma_1 = \cdots = \Gamma_G.$

In M_1, $\mathbf{\Lambda}$ is invariant over the two groups, while in M_2, $\mathbf{\Gamma}$ is invariant over the groups. 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:

$$\begin{aligned} M_{ta1}: \quad & \mathbf{v}_i^{(g)} = \boldsymbol{\mu}^{(g)} + t\mathbf{\Lambda}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}, \\ & \underline{\boldsymbol{\eta}_i^{(g)} = \mathbf{\Gamma}^{(g)}\mathbf{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\delta}_i^{(g)}, \quad g = 1, 2, \ i = 1, \cdots, N_g.} \end{aligned}$$

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 $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$, $\mathbf{\Lambda}$, $\mathbf{\Psi}_\epsilon^{(1)}$, $\mathbf{\Psi}_\epsilon^{(2)}$, $\mathbf{\Gamma}^{(1)}$, $\mathbf{\Gamma}^{(2)}$, $\mathbf{\Phi}^{(1)}$, $\mathbf{\Phi}^{(2)}$, $\mathbf{\Psi}_\delta^{(1)}$, and $\mathbf{\Psi}_\delta^{(2)}$.

The link model M_{ta2} for linking M_2 and M_a is defined as follows:

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

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 the next section.

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. We use a synthetic two-sample data that are randomly drawn from two populations (China and western countries). The sample sizes are $N_1 = 338$ and $N_2 = 247$.

Twenty-six ordered categorical items out of 100 original items are taken from the WHOQOL-100 instrument. They measure the 'overall QOL', 'physical health', 'mental health', 'social relationships', and 'environment', respectively. All of the items are measured with a 5-point scale. Hence, we consider a two-sample SEM with ordered categorical data. The ordered categorical variables are identified through 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 $\boldsymbol{\omega}_i^{(g)T} = (\eta_i^{(g)}, \xi_{i1}^{(g)}, \xi_{i2}^{(g)}, \xi_{i3}^{(g)}, \xi_{i4}^{(g)})$ are '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

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

with $\boldsymbol{\Lambda}^{(g)}$ 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 $\boldsymbol{\Lambda}_{0k}^{(g)}$ are taken as 0.8, elements in $\boldsymbol{\Lambda}_{0\omega k}^{(g)}$ 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;

$H_2 : \ \mathbf{\Lambda}^{(1)} = \mathbf{\Lambda}^{(2)}$;

$H_3 : \ \mathbf{\Lambda}^{(1)} = \mathbf{\Lambda}^{(2)}, \ \mathbf{\Phi}^{(1)} = \mathbf{\Phi}^{(2)}$.

The software WinBUGS was applied to obtain the Bayesian results. In the analysis:

- The number of burn-in iterations was taken as 10,000.
- After convergence, additional 10,000 observations were collected to produce the results.
- The DIC values corresponding to M_1, M_2, and M_3 are respectively equal to 32302.6, 32321.7, and 32341.9, suggesting M_1.
- The Bayesian estimates and their standard error estimates produced by WinBUGS under M_1 are presented in the following Table 6.4.

	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			

In this chapter, we have discussed

1. Multilevel SEMs, especially two-level SEMs. In analyzing these models, one should keep the following in mind:
    - The observations in hierarchical data are dependent.
    - In establishing a multi-level SEM, a simple between-group model is usually recommended.
    - Two sets of unknown parameters and latent variables that are associated with the within-group (first level) and between-group (second level) are considered.
    - The parameter estimation in between-group model is less accurate than those in within-group model.
    - Model comparison is non-trivial when comparing models with different hierarchical strucutures.
2. Multisample SEMs. In analyzing these models, one should note that
    - Differ multisample data from hierarchical and heterogeneous data.
    - Understand the main purpose of the multisample analysis.
    - Specify the prior distributions and derive the associated posterior distributions for constrained unknown parameters.
    - Convert the hypothesis testing problem to a model comparison problem.