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 $\boldsymbol{\Psi}_2$, α_{02k}, β_{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