Basics of SEM

· **Standard Assumptions** (1) $\boldsymbol{\epsilon}_i$ i.i.d. according to $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$, where $\boldsymbol{\Psi}_\epsilon$ is diagonal. (2) $\boldsymbol{\xi}_i$ are i.i.d. according to $N[\mathbf{0}, \boldsymbol{\Phi}]$, where $\boldsymbol{\Phi}$ is a general covariance matrix. (3) $\boldsymbol{\delta}_i$ are i.i.d. according to $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$, where $\boldsymbol{\Psi}_\delta$ is diagonal. (4) $\boldsymbol{\delta}_i$ is independent of $\boldsymbol{\xi}_i$, and $\boldsymbol{\epsilon}_i$ is independent of $\boldsymbol{\omega}_i$ and $\boldsymbol{\delta}_i$. Formula : $\boldsymbol{\eta}_i = \mathbf{B}\mathbf{d}_i + \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i = \boldsymbol{\Lambda}_\omega \mathbf{G}(\boldsymbol{\omega}_i) + \boldsymbol{\delta}_i$, $\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$.

· **Identifiability** The measurement equation as identified if for any $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2, m(\boldsymbol{\theta}_1) = m(\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.The structural equation as identified if for any $\boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_2^*, s(\boldsymbol{\theta}_1^*) = s(\boldsymbol{\theta}_2^*)$ implies $\boldsymbol{\theta}_1^* = \boldsymbol{\theta}_2^*$. The SEM as identified if both of its measurement equation and structural equation are identified. (1) Using a Λ with the non-overlapping structure. (2) fixing the diagonal elements of $\boldsymbol{\Phi}^+$ (covariance matrix of $\boldsymbol{\omega}$) as 1 to restricts the variances of latent variables to be 1 (hence $\boldsymbol{\Phi}^+$is a correlation matrix).

· Inverted Gamma distribution : $\theta \stackrel{D}{=}$ Inverted Gamma$[\alpha, \beta]$ $p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta}$, $E(\theta) = \frac{\beta}{\alpha-1}$, $\mathrm{Var}(\theta) = \frac{\beta^2}{(\alpha-1)^2(\alpha-1)}$

· Inverted Wishart distribution : $\mathbf{W} \stackrel{D}{=} IW_q[\mathbf{R}_0^{-1}, \rho_0]$ $p(\mathbf{W}) = [2^{\rho_0 q/2}\pi^{q(q-1)/4}\prod_{i=1}^q \Gamma(\frac{\rho_0+1-i}{2})]^{-1} \cdot |\mathbf{R}_0|^{-\rho_0/2}$. $|\mathbf{W}|^{-(\rho_0+q+1)/2} \cdot \exp\{-\frac{1}{2}\operatorname{tr}(\mathbf{R}_0^{-1}\mathbf{W}^{-1})\}$, $E(\mathbf{W}) = \frac{\mathbf{R}_0^{-1}}{\rho_0-q-1}$. $\mathbf{W}^{-1} \stackrel{D}{=} W_q[\mathbf{R}_0, \rho_0]$, $E(\mathbf{W}^{-1}) = \rho_0\mathbf{R}_0$.

Bayesian Estimating of SEM

· **Prior :** (1) $\psi_{\epsilon k}^{-1} \stackrel{D}{=}$ Gamma$[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$, $\boldsymbol{\mu} \stackrel{D}{=} N[\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0]$ and $[\boldsymbol{\Lambda}_k|\psi_{\epsilon k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}]$, where $\boldsymbol{\Sigma}_0$ and $\mathbf{H}_{0yk}$ are positive definite. (2) $\boldsymbol{\Phi}^{-1} \stackrel{D}{=} W_{q_2}[\mathbf{R}_0, \rho_0]$, $\psi_{\delta k}^{-1} \stackrel{D}{=}$ Gamma$[\alpha_{0\delta k}, \beta_{0\delta k}]$, and $[\boldsymbol{\Lambda}_{\omega k}|\psi_{\delta k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0\omega k}, \psi_{\delta k}\mathbf{H}_{0\omega k}]$, where $\mathbf{R}_0$ and $\mathbf{H}_{0\omega k}$ are positive definite.

· **Posterior :** (1) $p(\boldsymbol{\omega}_i \mid \mathbf{y}_i, \boldsymbol{\theta}) \propto p(\mathbf{y}_i \mid \boldsymbol{\omega}_i, \boldsymbol{\theta})p(\boldsymbol{\omega}_i \mid \boldsymbol{\theta})$ $\propto \exp\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T\boldsymbol{\Psi}_\epsilon^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) - \frac{1}{2}(\boldsymbol{\omega}_i - \boldsymbol{\mu}_\omega)^T\boldsymbol{\Sigma}_\omega^{-1}(\boldsymbol{\omega}_i - \boldsymbol{\mu}_\omega)\}$ $\propto \exp\{-\frac{1}{2}[\mathbf{y}_i^T\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i - 2\boldsymbol{\omega}_i^T\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i + \boldsymbol{\omega}_i^T(\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda})\boldsymbol{\omega}_i +$ $\boldsymbol{\omega}_i^T\boldsymbol{\Sigma}_\omega^{-1}\boldsymbol{\omega}_i - 2\boldsymbol{\omega}_i^T\boldsymbol{\Sigma}_\omega^{-1}\boldsymbol{\mu}_\omega\} \propto \exp\{-\frac{1}{2}[\boldsymbol{\omega}_i - \boldsymbol{\Sigma}^{*^{-1}}(\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i +$ $\boldsymbol{\Sigma}_\omega^{-1}\boldsymbol{\mu}_\omega)]^T\boldsymbol{\Sigma}^*[\boldsymbol{\omega}_i - \boldsymbol{\Sigma}^{*^{-1}}(\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i + \boldsymbol{\Sigma}_\omega^{-1}\boldsymbol{\mu}_\omega)\}$ Thus, $[\boldsymbol{\omega}_i \mid \mathbf{y}_i, \boldsymbol{\theta}] \stackrel{D}{=} N[\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i + \boldsymbol{\Sigma}^{*^{-1}}\boldsymbol{\Sigma}_\omega^{-1}\boldsymbol{\mu}_\omega, \boldsymbol{\Sigma}^{*-1}]$ where $\boldsymbol{\Pi}_0 = \mathbf{I} - \boldsymbol{\Pi}$, $\boldsymbol{\mu}_\omega = ((\boldsymbol{\Pi}_0^{-1}\mathbf{B}\mathbf{d}_i)^T, \mathbf{0}^T)^T$, $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_\omega^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda}$, and

$$\boldsymbol{\Sigma}_\omega = \begin{bmatrix} \boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^T + \boldsymbol{\Psi}_\delta)\boldsymbol{\Pi}_0^{-T} & \boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}^T\boldsymbol{\Pi}_0^{-T} & \boldsymbol{\Phi} \end{bmatrix}$$

(2) Let $\nu_k = \psi_{\epsilon k}^{-1}$. $p(\nu_k) \propto \nu_k^{\alpha_{0k}-1}\exp(-\beta_{0k}\nu_k)$.
$p(\boldsymbol{\Lambda}_k|\nu_k) \propto \nu_k^{q/2}\exp[-\frac{1}{2}(\boldsymbol{\Lambda}_k - \boldsymbol{\Lambda}_{0k})^T\mathbf{H}_{0yk}^{-1}(\boldsymbol{\Lambda}_k - \boldsymbol{\Lambda}_{0ykk})\,\nu_k]$.
$p(\mathbf{Y}|\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\Omega}) \propto |\boldsymbol{\Psi}_\epsilon|^{-n/2}\exp[-\frac{1}{2}\sum_{i=1}^n(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T\boldsymbol{\Psi}_\epsilon^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)]$.
Let $\mathbf{Y}_k^T$ be the k th row of $\mathbf{Y}, y_{ik}$ be the i th component of $\mathbf{Y}_k^T, \mathbf{A}_k^* =$ $(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)^{-1}\boldsymbol{\Omega}\mathbf{Y}_k$, and $b_k = \mathbf{Y}_k^T\mathbf{Y}_k - \mathbf{Y}_k^T\boldsymbol{\Omega}^T(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)^{-1}\boldsymbol{\Omega}\mathbf{Y}_k = \mathbf{Y}_k^T\mathbf{Y}_k -$ $\mathbf{A}_k^{*T}(\boldsymbol{\Omega}^T)\mathbf{A}_k^*$. The exponential term in $p(\mathbf{Y}|\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\Omega})$ is $-\frac{1}{2}\sum_{i=1}^n(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T\boldsymbol{\Psi}_\epsilon^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) = -\frac{1}{2}\sum_{i=1}^n\sum_{k=1}^p\psi_{\epsilon k}^{-1}(y_{ik} -$ $\boldsymbol{\Lambda}_k^T\boldsymbol{\omega}_i)^2 = -\frac{1}{2}\sum_{k=1}^p\{\nu_k[\sum_{i=1}^n y_{ik}^2 - 2\boldsymbol{\Lambda}_k^T\sum_{i=1}^n y_{ki}\boldsymbol{\omega}_i + \operatorname{tr}(\boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k^T$ $\sum_{i=1}^n\boldsymbol{\omega}_i\boldsymbol{\omega}_i^T)]\} = -\frac{1}{2}\sum_{k=1}^p\{\nu_k[\mathbf{Y}_k^T\mathbf{Y}_k - 2\boldsymbol{\Lambda}_k^T\boldsymbol{\Omega}\mathbf{Y}_k + \boldsymbol{\Lambda}_k^T(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)\boldsymbol{\Lambda}_k]\} =$ $-\frac{1}{2}\sum_{k=1}^p\{\nu_k[b_k + (\boldsymbol{\Lambda}_k - \mathbf{A}_k^*)^T(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)(\boldsymbol{\Lambda}_k - \mathbf{A}_k^*)]\}$.
$p(\boldsymbol{\Lambda}, \nu_1, \ldots, \nu_p|\mathbf{Y}, \boldsymbol{\Omega}) \propto \prod_{k=1}^p[\nu_k^{n/2+q/2+\alpha_{0\epsilon k}-1}\exp\{-\frac{1}{2}\nu_k$ $[(\boldsymbol{\Lambda}_k - \mathbf{A}_k^*)^T(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)(\boldsymbol{\Lambda}_k - \mathbf{A}_k^*) + (\boldsymbol{\Lambda}_k - \boldsymbol{\Lambda}_{0k})^T\mathbf{H}_{0yk}^{-1}(\boldsymbol{\Lambda}_k - \boldsymbol{\Lambda}_{0k})] -$ $\nu_k(\beta_{0\epsilon k} + b_k/2)\}] = \prod_{k=1}^p p(\boldsymbol{\Lambda}_k, \nu_k|\mathbf{Y}, \boldsymbol{\Omega})$.
Let $\mathbf{A}_k = (\mathbf{H}_{0yk}^{-1} + \boldsymbol{\Omega}\,\boldsymbol{\Omega}^T)^{-1}$ and $\mathbf{a}_k = \mathbf{A}_k(\mathbf{H}_{0yk}^{-1}\boldsymbol{\Lambda}_{0k} + \boldsymbol{\Omega}\mathbf{Y}_k)$, then $(\boldsymbol{\Lambda}_k - \mathbf{A}_k^*)^T(\boldsymbol{\Omega}\boldsymbol{\Omega}^T)(\boldsymbol{\Lambda}_k - \mathbf{A}_k^*) + (\boldsymbol{\Lambda}_k - \boldsymbol{\Lambda}_{0k})^T\mathbf{H}_{0yk}^{-1}(\boldsymbol{\Lambda}_k - \boldsymbol{\Lambda}_{0k}) =$ $(\boldsymbol{\Lambda}_k - \mathbf{a}_k)^T\mathbf{A}_k^{-1}(\boldsymbol{\Lambda}_k - \mathbf{a}_k) - \mathbf{a}_k^T\mathbf{A}_k^{-1}\mathbf{a}_k + \mathbf{A}_k^{*T}\boldsymbol{\Omega}\boldsymbol{\Omega}^T\mathbf{A}_k^* + \boldsymbol{\Lambda}_{0k}^T\mathbf{H}_{0yk}^{-1}\boldsymbol{\Lambda}_{0k}$. Hence $p(\boldsymbol{\Lambda}_k, \nu_k|\mathbf{Y}, \boldsymbol{\Omega}) = p(\nu_k|\mathbf{Y}, \boldsymbol{\Omega})p(\boldsymbol{\Lambda}_k|\mathbf{Y}, \boldsymbol{\Omega}, \nu_k) \propto [\nu_k^{n/2+\alpha_{0\epsilon k}-1}$ $\exp(-\beta_{\epsilon k}\nu_k)] \cdot \{\nu_k^{q/2}\exp[-\frac{1}{2}(\boldsymbol{\Lambda}_k - \mathbf{a}_k)^T\mathbf{A}_k^{-1}(\boldsymbol{\Lambda}_k - \mathbf{a}_k)\nu_k]\}$ where $\beta_{\epsilon k} = \beta_{0\epsilon k} + 2^{-1}(\mathbf{Y}_k^T\mathbf{Y}_k - \mathbf{a}_k^T\mathbf{A}_k^{-1}\mathbf{a}_k + \boldsymbol{\Lambda}_{0k}^T\mathbf{H}_{0yk}^{-1}\boldsymbol{\Lambda}_{0k})$. Thus, $[\nu_k|\mathbf{Y}, \boldsymbol{\Omega}] \stackrel{D}{=}$ Gamma$[n/2+\alpha_{0\epsilon k}, \beta_{\epsilon k}]$, and $[\boldsymbol{\Lambda}_k|\mathbf{Y}, \boldsymbol{\Omega}, \nu_k] \stackrel{D}{=} N[\mathbf{a}_k, \nu_k^{-1}\mathbf{A}_k]$. (3) $p(\boldsymbol{\Phi} \mid \boldsymbol{\Omega}_2) \propto p(\boldsymbol{\Phi})\prod_{i=1}^n p(\boldsymbol{\xi}_i \mid \boldsymbol{\theta})$. Then $p(\boldsymbol{\Phi} \mid \boldsymbol{\Omega}_2) \propto [|\boldsymbol{\Phi}|^{-(\rho_0+q_2+1)/2}\exp\{-\frac{1}{2}\operatorname{tr}[\mathbf{R}_0^{-1}\boldsymbol{\Phi}^{-1}]\}]\cdot$ $[|\boldsymbol{\Phi}|^{-n/2}\exp\{-\frac{1}{2}\sum_{i=1}^n\boldsymbol{\xi}_i^T\boldsymbol{\Phi}^{-1}\boldsymbol{\xi}_i\}]$ $= |\boldsymbol{\Phi}|^{-(n+\rho_0+q_2+1)/2}\exp\{-\frac{1}{2}\operatorname{tr}[\boldsymbol{\Phi}^{-1}(\boldsymbol{\Omega}_2\boldsymbol{\Omega}_2^T + \mathbf{R}_0^{-1})]\}$. Hence $[\boldsymbol{\Phi} \mid \boldsymbol{\Omega}_2] \stackrel{D}{=} IW_{q_2}[(\boldsymbol{\Omega}_2\boldsymbol{\Omega}_2^T + \mathbf{R}_0^{-1}), n + \rho_0]$.

· If some elements of $\boldsymbol{\Lambda}_k$ are fixed, we identify the positions of the fixed elements via an index matrix $\mathbf{L}$ with the following elements : $I_{kj} = \{ \begin{smallmatrix} 0, & \text{if } \lambda_{kj} \text{ is fixed,} \\ 1, & \text{if } \lambda_{kj} \text{ is free;} \end{smallmatrix}$ for $j = 1, \cdots, q$ and $k = 1, \cdots, p$.. Let $\boldsymbol{\Lambda}_k^*$ be a vector of unknown parameters in $\boldsymbol{\Lambda}_k, \mathbf{Y}_k$ be the submatrix of $\mathbf{Y}$ such that all the rows corresponding to $I_{kj} = 0$ are deleted ; and let $\mathbf{Y}_k^{*T} = (y_{1k}^*, \cdots, y_{nk}^*)$ with $y_{ik}^* = y_{ik} - \sum_{j=1}^q\lambda_{kj}y_{ij}(1 - l_{kj})$ where y_{ij} is the j-th element of $\mathbf{y}_i$. Then, $[\nu_k|\mathbf{Y}, \boldsymbol{\Omega}] \stackrel{D}{=}$ Gamma$[n/2 + \alpha_{0\epsilon k}, \beta_{\epsilon k}]$, $[\boldsymbol{\Lambda}_k^*|\mathbf{Y}, \boldsymbol{\Omega}, \nu_k] \stackrel{D}{=} N[\mathbf{a}_k, \nu_k\mathbf{A}_k]$, where $\mathbf{A}_k = (\mathbf{H}_{0yk}^{-1} + \mathbf{Y}_k\mathbf{Y}_k^T)^{-1}$, $\mathbf{a}_k = \mathbf{A}_k(\mathbf{H}_{0yk}^{-1}\boldsymbol{\Lambda}_{0yk} + \boldsymbol{\Omega}\mathbf{Y}_k^*)$, and $\beta_{\epsilon k} = \beta_{0\epsilon k} + \frac{1}{2}(\mathbf{Y}_k^{*T}\mathbf{Y}_k^* - \mathbf{a}_k^T\mathbf{A}_k^{-1}\mathbf{a}_k$ $+\boldsymbol{\Lambda}_{0k}^T\mathbf{H}_{0vk}^{-1}\boldsymbol{\Lambda}_{0k})$.

· **Convergence :** (1) At convergence, parallel sequences generated with different starting values should mix well together. (2) Using estimated potential scale reduction (EPSR) value. Convergence is achieved when the EPSR values are all less than 1.2. $B = \frac{n}{K-1}\sum_{k=1}^K(\theta_{\cdot k} - \theta_{\cdot\cdot})^2$, $\theta_{\cdot k} = n^{-1}\sum_{j=1}^n\theta_{jk}$, $\theta_{\cdot\cdot} = K^{-1}\sum_{k=1}^K\theta_k$, $W = \frac{1}{K}\sum_{k=1}^K s_k^2$, $s_k^2 = (n-1)^{-1}\sum_{j=1}^n(\theta_{jk} - \theta_{\cdot k})^2$. $\widehat{\mathrm{var}}(\theta) = \frac{n-1}{n}W +$ $\frac{1}{n}B$. The EPSR is defined as $\hat{R}^{1/2} = [\widehat{\mathrm{var}}(\theta)/W]^{1/2}$.

Bayes Factor

· Let $p(M_0)$ be the prior probability of M_0 and $p(M_1) = 1 - p(M_0)$, and let $p(M_k|\mathbf{Y})$ be the posterior probability for $k = 0, 1$. From the Bayes theorem, we have $p(M_k|\mathbf{Y}) = \frac{p(\mathbf{Y}|M_k)p(M_k)}{p(\mathbf{Y}|M_1)p(M_1)+p(\mathbf{Y}|M_0)p(M_0)}$, $k = 0, 1$. Hence $\frac{p(M_1|\mathbf{Y})}{p(M_0|\mathbf{Y})} = \frac{p(\mathbf{Y}|M_1)p(M_1)}{p(\mathbf{Y}|M_0)p(M_0)}$. The Bayes factor for comparing M_1 and M_0 is defined as $B_{10} = \frac{p(\mathbf{Y}|M_1)}{p(\mathbf{Y}|M_0)}$. **Note** (1) It may reject a null hypothesis associated with M_0, or may equally provide evidence in favor of the null hypothesis or the alternative hypothesis associated with M_1. (2) The comparison based on the Bayes factor does not depend on the assumption that either model is 'true'. (3) The same data set is used in the comparison ; hence, it does not favor the alternative hypothesis (or M_1) in extremely large samples. (4) It can be applied to compare nonnested models M_0 and M_1.

B_{10}	$2\log B_{10}$	Evidence against $H_0(M_0)$
< 1	< 0	Negative (supports $.H_0(M_0)$)
1 to 3	0 to 2	Not worth more than a bare mention
3 to 20	2 to 6	Positive (supports $.H_1(M_1)$)
20 to 150	6 to 10	Strong
> 150	> 10	Decisive

· (Path Sampling) $p(\mathbf{Y}|M_k) = \int p(\mathbf{Y}|\boldsymbol{\theta}_k, M_k)p(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k$ is difficult to obtain B_{10} analytically. Consider a class of densities which are denoted by a continuous parameter t in $[0, 1]$: $p(\boldsymbol{\Omega}, \boldsymbol{\theta}|\mathbf{Y}, t) = \frac{1}{z(t)}\,p(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}|t)$, where $z(t) = p(\mathbf{Y}|t) = \int p(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}|t)d\boldsymbol{\Omega}d\boldsymbol{\theta} = \int p(\mathbf{Y}, \boldsymbol{\Omega}, |\boldsymbol{\theta}, t)p(\boldsymbol{\theta})d\boldsymbol{\Omega}d\boldsymbol{\theta}$. We construct a path using the parameter t in $[0, 1]$ to link two competing models M_1 and M_0 together, so that $z(1) =$ $p(\mathbf{Y}|1) = p(\mathbf{Y}|M_1)$, $z(0) = p(\mathbf{Y}|0) = p(\mathbf{Y}|M_0)$, and $B_{10} = z(1)/z(0)$. Taking logarithm and then differentiating $z(t)$ with respect to t, and assuming the legitimacy of interchange of integration with differentiation, we have $\frac{d\log z(t)}{dt} = \int\frac{1}{z(t)}\frac{d}{dt}p(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}|t)d\boldsymbol{\Omega}d\boldsymbol{\theta} =$ $\int\frac{d}{dt}\log p(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}|t)\cdot p(\boldsymbol{\Omega}, \boldsymbol{\theta}|\mathbf{Y}, t)d\boldsymbol{\Omega}d\boldsymbol{\theta} = E_{\Omega,\theta}[\frac{d}{dt}\log p(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}|t)]$, where $E_{\Omega,\theta}$ denotes the expectation with respect to the distribution $p(\boldsymbol{\Omega}, \boldsymbol{\theta}|\mathbf{Y}, t)$. Let $U(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}, t) = \frac{d}{dt}\log p(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}|t) = \frac{d}{dt}\log p(\mathbf{Y}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t)$ which does not involve the prior density $p(\boldsymbol{\theta})$, we have $\log B_{10} =$ $\log\frac{z(1)}{z(0)} = \int_0^1 E_{\Omega,\theta}[U(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}, t)]dt$. We first order the unique values of fixed grids $\{t_{(s)}\}_{s=1}^S$ between $[0, 1]$ such that $0 = t_{(0)} < t_{(1)} < \cdots$ $< t_{(S)} < t_{(S+1)} = 1$, and estimate $\log B_{10}$ by $\widehat{\log B_{10}} = \frac{1}{2}\sum_{s=0}^S$ $(t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_{(s)})$ where $\bar{U}_{(s)}$ is the following average of the values of $U(\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}, t)$ based on simulation draws at $t = t_{(s)}$ $\bar{U}_{(s)} = J^{-1}\sum_{j=1}^J U(\mathbf{Y}, \boldsymbol{\Omega}^{(j)}, \boldsymbol{\theta}^{(j)}, t_{(s)})$ in which $\{(\boldsymbol{\Omega}^{(j)}, \boldsymbol{\theta}^{(j)}),$ $j = 1, \cdots, J\}$ are observations drawn from $p(\boldsymbol{\Omega}, \boldsymbol{\theta}|\mathbf{Y}, t_{(s)})$.

Other Model Comparison Statistics

· (BIC) An approximation of $2\log B_{10}$ that does not depend on the prior density is $2\log B_{10} \cong 2S^* = 2\{\log p(\mathbf{Y}|\tilde{\boldsymbol{\theta}}_1, M_1) - \log p(\mathbf{Y}|\tilde{\boldsymbol{\theta}}_0,$ $M_0)\} - (d_1 - d_0)\log n$ where $\tilde{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_0$ are the maximum likelihood (ML) estimates of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_0$ under M_1 and M_0, respectively ; d_1 and d_0 are the dimensions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_0$, and n is the sample size. Minus $2S^*$ is the following well-known Bayesian Information Criterion (BIC) for comparing M_1 and M_0 : $\mathrm{BIC}_{10} = -2S^* \cong -2\log B_{10} = 2\log B_{01}$. Alternatively, for each M_k, $k = 0, 1$, we can define $\mathrm{BIC}_k = -2\log p(\mathbf{Y}|$ $\tilde{\boldsymbol{\theta}}_k, M_k) + d_k\log n$. Hence $2\log B_{10} \cong \mathrm{BIC}_0 - \mathrm{BIC}_1$ and the smaller BIC_k value is selected.

· (AIC) The Akaike Information Criterion (AIC ; Akaike, 1973) associated with a competing model M_k is given by $\mathrm{AIC}_k = -2\log p(\mathbf{Y}|$ $\tilde{\boldsymbol{\theta}}_k, M_k) + 2d_k$ which does not involve the sample size n. We see that BIC tends to favor simpler models.

· (DIC) Under a competing model M_k with a vector of unknown parameter $\boldsymbol{\theta}_k$, the DIC is defined as $\mathrm{DIC}_k = \overline{D(\boldsymbol{\theta}_k)} + d_k$ where $\overline{D(\boldsymbol{\theta}_k)}$ measures the goodness-of-fit of the model, and is defined as $\overline{D(\boldsymbol{\theta}_k)} =$ $E_{\boldsymbol{\theta}_k}\{-2\log p(\mathbf{Y}|\boldsymbol{\theta}_k, M_k)|\mathbf{Y}\}$. Here, d_k is the effective number of parameters in M_k, and is defined as $d_k = E_{\boldsymbol{\theta}_k}\{-2\log p(\mathbf{Y}|\boldsymbol{\theta}_k, M_k)|$ $\mathbf{Y}\} + 2\log p(\mathbf{Y}|\tilde{\boldsymbol{\theta}}_k)$ in which $\tilde{\boldsymbol{\theta}}_k$ is the Bayesian estimate of $\boldsymbol{\theta}_k$.

· (L_ν-Measure) It measures the performance of a model by a combination of how close its predictions are to the observed data and the variability of the predictions. Let $\mathbf{Y}$ be the observed data, and let $p(\mathbf{Y}, \boldsymbol{\theta})$ be the joint density that corresponds to a model M with a parameter vector $\boldsymbol{\theta}$. The future responses $\mathbf{Y}^{\text{rep}} = (\mathbf{y}_1^{\text{rep}}, \cdots, \mathbf{y}_n^{\text{rep}})$, which have the same sampling density as $p(\mathbf{Y}|\boldsymbol{\theta})$. For some $\delta > 0$, let $L_1(\mathbf{Y}, \mathbf{B}, \delta) = E[\operatorname{tr}(\mathbf{Y}^{\text{rep}} - \mathbf{B})^T(\mathbf{Y}^{\text{rep}} - \mathbf{B})] + \delta\operatorname{tr}(\mathbf{Y} - \mathbf{B})^T(\mathbf{Y} - \mathbf{B})$ where the expectation is taken with respect to the posterior predictive distribution of $[\mathbf{Y}^{\text{rep}}|\mathbf{Y}]$. Note that this statistic reduces to the Euclidean distance by setting $\mathbf{B} = \mathbf{Y}$. By setting $\mathbf{B}$ as the minimizer, it can be shown that $L_\nu(\mathbf{Y}) = \sum_{i=1}^n\operatorname{tr}\{\mathrm{Cov}(\mathbf{y}_i^{\text{rep}}|\mathbf{Y})\} + \nu\sum_{i=1}^n\operatorname{tr}[\{E($ $\mathbf{y}_i^{\text{rep}}|\mathbf{Y}) - \mathbf{y}_i\}\{E(\mathbf{y}_i^{\text{rep}}|\mathbf{Y}) - \mathbf{y}_i\}^T]$ where $\nu = \delta/(\delta + 1)$. This statistic is called the L_ν-measure.

Ordered Categorical Data

· Consider the measurement equation for a $p \times 1$ observed random vector $\mathbf{v}_i : \mathbf{v}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$, $i = 1, \cdots, n$ where $\mathbf{v} = (\mathbf{x}^T, \mathbf{y}^T)^T$, where $\mathbf{x}$ continuous measurements are observable, $\mathbf{y} = (y_1, \cdots, y_s)^T$ is the subset of unobservable continuous measurements. The information associated with $\mathbf{y}$ is given by observable ordered categorical vector $\mathbf{z}$.

· **Identifiability :** (1) Fixing appropriate elements in $\boldsymbol{\Lambda}, \boldsymbol{\Pi}$, and/or $\boldsymbol{\Gamma}$

$\log p(Y,\Omega|\theta,t) = \sum_i \log\{\sum_{c=1}^{C}[\pi_c + (1-t)a_c \sum_{k=c+1}^{r}\pi_k] f_k(y_i, w_i|\mu_k,\Sigma_k) + \sum_{k=c+1}^{r} t\pi_k f_k(y_i,w_i|\mu_k,\Sigma_k)\}$

M_t: $\pi_1 + (1-t)a_1(\pi_{c+1}+\cdots+\pi_k)] f_1(y|\mu_1,\Sigma_1) + \cdots + (\pi_c + (1-t)a_c(\pi_{c+1}+\cdots+\pi_k)] f_c(y_c|\mu_c,\Sigma_c) + t\pi_{c+1}f_{c+1}(y|\mu_{c+1},\Sigma_{c+1}) + \cdots + t\pi_k f_k(y|\mu_k,\Sigma_k)$

at preassigned values. (2) For every k, we may fix $\alpha_{k,1} = \Phi^{*-1}(f^*_{k,1})$ and $\alpha_{k,b_k} = \Phi^{*-1}(f^*_{k,b_k})$, where $\Phi^*(\cdot)$ is the distribution function of $N[0,1]$, $f^*_{k,1}$ and f^*_{k,b_k} are the frequency of the first category, and the cumulative frequency of the category with $z_k < b_k$, respectively.

· **Bayes Analysis :** Let $\mathbf{X}$, $\mathbf{Z}$ be the observed continuous and ordered categorical, $\mathbf{Y}$ and $\mathbf{\Omega}$ be latent continuous and latent variables. The observed data $[\mathbf{X},\mathbf{Z}]$ are augmented with the latent data $[\mathbf{Y},\mathbf{\Omega}]$. To implement the Gibbs sampler, we start with initial starting values $(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\theta}^{(0)}, \mathbf{\Omega}^{(0)}, \mathbf{Y}^{(0)})$, then simulate $\mathbf{\Omega}^{(j+1)}$ from $p(\mathbf{\Omega}|\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{X}, \mathbf{Z})$, $\boldsymbol{\theta}^{(j+1)}$ from $p(\boldsymbol{\theta}|\mathbf{\Omega}^{(j+1)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{X}, \mathbf{Z})$, $(\boldsymbol{\alpha}^{(j+1)}, \mathbf{Y}^{(j+1)})$ from $p(\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}^{(j+1)}, \mathbf{\Omega}^{(j+1)}, \mathbf{X}, \mathbf{Z})$.

· **Conditional Distributions :** $p(\mathbf{\Omega}|\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}) = \prod_{i=1}^n p(\boldsymbol{\omega}_i|\mathbf{v}_i, \boldsymbol{\theta}) \propto \prod_{i=1}^n \exp\{-\frac{1}{2}[\boldsymbol{\xi}_i^T\mathbf{\Phi}^{-1}\boldsymbol{\xi}_i + (\mathbf{v}_i - \boldsymbol{\mu} - \mathbf{\Lambda}\boldsymbol{\omega}_i)^T\mathbf{\Psi}_\epsilon^{-1}(\mathbf{v}_i - \boldsymbol{\mu} - \mathbf{\Lambda}\boldsymbol{\omega}_i) + (\boldsymbol{\eta}_i - \mathbf{\Lambda}_\omega\boldsymbol{\omega}_i)^T\mathbf{\Psi}_\delta^{-1}(\boldsymbol{\eta}_i - \mathbf{\Lambda}_\omega\boldsymbol{\omega}_i)]\}$. Let the prior to be $\boldsymbol{\mu} \stackrel{D}{=} N[\boldsymbol{\mu}_0, \mathbf{\Sigma}_0]$, $\psi_{\epsilon k}^{-1} \stackrel{D}{=} \text{Gamma}[\alpha_{0\epsilon k}, \beta_{0\epsilon k}]$, $[\mathbf{\Lambda}_k|\psi_{\epsilon k}] \stackrel{D}{=} N[\mathbf{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0vk}]$, $\psi_{\delta k}^{-1} \stackrel{D}{=} \text{Gamma}[\alpha_{0\delta k}, \beta_{0\delta k}]$, $[\mathbf{\Lambda}_{\omega k}|\psi_{\delta k}] \stackrel{D}{=} N[\mathbf{\Lambda}_{0\omega k}, \psi_{\delta k}\ \mathbf{H}_{0\omega k}]$. Let $\mathbf{A}_k = (\mathbf{H}_{0vk}^{-1} + \mathbf{\Omega}_k\mathbf{\Omega}_k^T)^{-1}$, $\mathbf{a}_k = \mathbf{A}_k(\mathbf{H}_{0vk}^{-1}\mathbf{\Lambda}_{0k} + \mathbf{\Omega}_k\mathbf{V}_k^*)$ and $\beta_{\epsilon k} = \beta_{0\epsilon k} + 2^{-1}(\mathbf{V}_k^{*T}\mathbf{V}_k^* - \mathbf{a}_k^T\mathbf{A}_k^{-1}\mathbf{a}_k + \mathbf{\Lambda}_{0k}^T\mathbf{H}_{0vk}^{-1}\mathbf{\Lambda}_{0k}$. Then $p(\psi_{\epsilon k}^{-1}|\boldsymbol{\mu}, \mathbf{V}, \mathbf{\Omega}) \stackrel{D}{=} \text{Gamma}[n/2 + \alpha_{0\epsilon k}, \beta_{\epsilon k}]$, $p(\mathbf{\Lambda}_k|\psi_{\epsilon k}^{-1}, \boldsymbol{\mu}, \mathbf{V}, \mathbf{\Omega}) \stackrel{D}{=} N[\mathbf{a}_k, \psi_{\epsilon k}\mathbf{A}_k]$, $p(\boldsymbol{\mu}|\mathbf{\Lambda}, \mathbf{\Psi}_\epsilon, \mathbf{V}, \mathbf{\Omega}) \stackrel{D}{=} N[(\mathbf{\Sigma}_0^{-1} + n\mathbf{\Psi}_\epsilon^{-1})^{-1}(n\mathbf{\Psi}_\epsilon^{-1}\overline{\mathbf{V}} + \mathbf{\Sigma}_0^{-1}\boldsymbol{\mu}_0), (\mathbf{\Sigma}_0^{-1} + n\mathbf{\Psi}_\epsilon^{-1})^{-1}]$, $p(\psi_{\delta k}^{-1}|\mathbf{\Omega}) \stackrel{D}{=} \text{Gamma}[n/2 + \alpha_{0\delta k}, \beta_{\delta k}]$, $p(\mathbf{\Lambda}_{\omega k}|\mathbf{\Omega}, \psi_{\delta k}^{-1}) \stackrel{D}{=} N[\mathbf{a}_{\omega k}, \psi_{\delta k}\mathbf{A}_{\omega k}]$, $p(\mathbf{\Phi}|\mathbf{\Omega}_{(2)}) \stackrel{D}{=} IW_{q2}[(\mathbf{\Omega}_{(2)}\mathbf{\Omega}_{(2)}^T + \mathbf{R}_0^{-1}), n + \rho_0]$, where $\mathbf{\Omega}_2 = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)$, $\overline{\mathbf{V}} = \sum_{i=1}^n(\mathbf{v}_i - \mathbf{\Lambda}\boldsymbol{\omega}_i)/n$. Let $p(\boldsymbol{\alpha}_k) \propto c$ (non-informative), we have $p(\boldsymbol{\alpha}_k|\mathbf{Z}_k, \boldsymbol{\theta}, \mathbf{\Omega}) \propto \prod_{i=1}^n\{\Phi^*[\psi_{yk}^{-1/2}(\alpha_{k,z_{ik}+1} - \mu_{yk} - \mathbf{\Lambda}_{ykk}^T\boldsymbol{\omega}_i)] - \Phi^*[\psi_{yk}^{-1/2}(\alpha_{k,z_{ik}} - \mu_{yk} - \mathbf{\Lambda}_{yk}^T\boldsymbol{\omega}_i)]\}$, $p(y_{ik}|\boldsymbol{\alpha}_k, \mathbf{Z}_k, \boldsymbol{\theta}, \mathbf{\Omega}) \stackrel{D}{=} N(\mu_{yk} + \mathbf{\Lambda}_{yk}^T\boldsymbol{\omega}_i, \psi_{yk})\ I_{(\alpha_{k,z_{ik}}, \alpha_{k,z_{ik}+1}]}(y_{ik})$, $p(\boldsymbol{\alpha}_k, \mathbf{Y}_k|\mathbf{Z}_k, \boldsymbol{\theta}, \mathbf{\Omega}) \propto \prod_{i=1}^n \phi[\psi_{yk}^{-1/2}(y_{ik} - \mu_{yk} - \mathbf{\Lambda}_{yk}^T\boldsymbol{\omega}_i)]\ I_{[\alpha_{k,z_{ik}}, \alpha_{k,z_{ik}+1}]}(y_{ik})$, where $\Phi^*(\cdot)$ denotes the standard normal cdf, $\phi(\cdot)$ is the standard normal density.

Dichotomous Variables

· **Identifiability :** Suppose that the exact measurement of $\mathbf{y}_i$ is not available and its information is given by an observed dichotomous vector $\mathbf{z}_i$ such that for $k = 1, \cdots, p$, $z_{ik} = 1$ if $y_{ik} > 0$ and $z_{ik} = 0$ otherwise. Then $\Pr(z_{ik} = 1|\boldsymbol{\omega}_i, \mu_k, \mathbf{\Lambda}_k, \psi_{\epsilon k}) = \Pr(y_{ik} > 0|\boldsymbol{\omega}_i, \mu_k, \mathbf{\Lambda}_k, \psi_{\epsilon k}) = \Phi^*\{(\mathbf{\Lambda}_k^T/\psi_{\epsilon k}^{1/2})\boldsymbol{\omega}_i + \mu_k/\psi_{\epsilon k}^{1/2}\}$. $C\mathbf{\Lambda}_k^T/(C\psi_{\epsilon k}^{1/2}) = \mathbf{\Lambda}_k^T/\psi_{\epsilon k}^{1/2}$ and $C\mu_k/(C\psi_{\epsilon k}^{1/2}) = \mu_k/\psi_{\epsilon k}^{1/2}$ for any positive constant C. We fix $\psi_{\epsilon k} = 1.0$. The measurement and structural equations are identified by fixing the approximate elements of Λ and Λ_ω at preassigned values.

· **Bayes Analysis :** Let $\mathbf{Z}$ be dichotomous variables, $\mathbf{Y}$ be the latent continuous measurements. We have $p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{\Omega}, \mathbf{Z}) = \prod_{i=1}^n p(\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\omega}_i, \mathbf{z}_i)$ where
$$[y_{ik}|\boldsymbol{\theta}, \boldsymbol{\omega}_i, \mathbf{z}_i] \stackrel{D}{=} \begin{cases} N[\mu_k + \mathbf{\Lambda}_k^T\boldsymbol{\omega}_i, 1]I_{(-\infty,0]}(y_{ik}), & \text{if } z_{ik} = 0, \\ N[\mu_k + \mathbf{\Lambda}_k^T\boldsymbol{\omega}_i, 1]I_{(0,\infty)}(y_{ik}), & \text{if } z_{ik} = 1. \end{cases}$$

Variables from Exponential Family Distributions

Consider $p(y_{ik}|\boldsymbol{\omega}_i) = \exp\{[y_{ik}\vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\epsilon k} + c_k(y_{ik}, \psi_{\epsilon k})\}$ $E(y_{ik}|\boldsymbol{\omega}_i) = \dot{b}(\vartheta_{ik})$, and $\text{Var}(y_{ik}|\boldsymbol{\omega}_i) = \psi_{\epsilon k}\ddot{b}(\vartheta_{ik})$ where $b(\cdot)$ and $c_k(\cdot)$ are specific differentiable functions with the dots denoting the derivatives. In addition, $\vartheta_{ik} = \mathbf{A}_k^T\mathbf{c}_{ik} + \mathbf{\Lambda}_k^T\boldsymbol{\omega}_i$.

· **Conditional Distributions :** $p(\mathbf{\Omega}|\mathbf{Y}, \boldsymbol{\theta}) = \prod_{i=1}^n p(\boldsymbol{\omega}_i|\mathbf{y}_i, \boldsymbol{\theta}) \propto \prod_{i=1}^n \exp\{\sum_{k=1}^p [y_{ik}\vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\epsilon k} - \frac{1}{2}[(\boldsymbol{\eta}_i - \mathbf{B}\mathbf{d}_i - \mathbf{\Pi}\boldsymbol{\eta}_i - \mathbf{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i))^T\mathbf{\Psi}_\delta^{-1}(\boldsymbol{\eta}_i - \mathbf{B}\mathbf{d}_i - \mathbf{\Pi}\boldsymbol{\eta}_i - \mathbf{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i)) + \boldsymbol{\xi}_i^T\mathbf{\Phi}^{-1}\boldsymbol{\xi}_i]\}$, $p(\mathbf{A}_k|\mathbf{Y}, \mathbf{\Omega}, \mathbf{\Lambda}_k, \psi_{\epsilon k}) \propto \exp\{\sum_{i=1}^n \frac{y_{ik}\vartheta_{ik} - b(\vartheta_{ik})}{\psi_{\epsilon k}} - \frac{1}{2}(\mathbf{A}_k - \mathbf{A}_{0k})^T\mathbf{H}_{0k}^{-1}(\mathbf{A}_k - \mathbf{A}_{0k})\}$, $p(\psi_{\epsilon k}|\mathbf{Y}, \mathbf{\Omega}, \mathbf{A}_k, \mathbf{\Lambda}_k) \propto \psi_{\epsilon k}^{-(\frac{n}{2} + \alpha_{0\epsilon k} - 1)}\exp\{\sum_{i=1}^n [\frac{y_{ik}\vartheta_{ik} - b(\vartheta_{ik})}{\psi_{\epsilon k}} + c_k(y_{ik}, \psi_{\epsilon k})] - \frac{\beta_{0k}}{\psi_{\epsilon k}}\}$, $p(\mathbf{\Lambda}_k|\mathbf{Y}, \mathbf{\Omega}, \mathbf{A}_k, \psi_{\epsilon k}) \propto \exp\{\sum_{i=1}^n \frac{y_{ik}\vartheta_{ik} - b(\vartheta_{ik})}{\psi_{\epsilon k}} - \frac{1}{2}\psi_{\epsilon k}^{-1}(\mathbf{\Lambda}_k - \mathbf{\Lambda}_{0k})^T\mathbf{H}_{0yk}^{-1}(\mathbf{\Lambda}_k - \mathbf{\Lambda}_{0k})\}$, $[\psi_{\delta k}^{-1}|\mathbf{\Omega}, \mathbf{\Lambda}_{\omega k}] \stackrel{D}{=} \text{Gamma}[n/2 + \alpha_{0\delta k}, \beta_{\delta k}]$, $[\mathbf{\Lambda}_{\omega k}|\mathbf{\Omega}, \psi_{\delta k}] \stackrel{D}{=} N[\boldsymbol{\mu}_{\omega k}, \psi_{\delta k}\mathbf{\Sigma}_{\omega k}]$, $[\mathbf{\Phi}|\mathbf{\Omega}] \stackrel{D}{=} IW_{q2}[(\mathbf{\Omega}_2\mathbf{\Omega}_2^T + \mathbf{R}_0^{-1}), n + \rho_0]$, where $\mathbf{\Sigma}_{\omega k} = (\mathbf{H}_{0\omega k}^{-1} + \mathbf{G}\mathbf{G}^T)^{-1}$, $\boldsymbol{\mu}_{\omega k} = \mathbf{\Sigma}_{\omega k}(\mathbf{H}_{0\omega k}^{-1}\mathbf{\Lambda}_{0\omega k} + \mathbf{G}\mathbf{\Omega}_{1k})$, and $\beta_{\delta k} = \beta_{0\delta k} + (\mathbf{\Omega}_{1k}^T\mathbf{\Omega}_{1k} - \boldsymbol{\mu}_{\omega k}^T\mathbf{\Sigma}_{\omega k}^{-1}\boldsymbol{\mu}_{\omega k} + \mathbf{\Lambda}_{0\omega k}^T\mathbf{H}_{0\omega k}^{-1}\mathbf{\Lambda}_{0\omega k})/2$, in which $\mathbf{G} = (\mathbf{G}(\boldsymbol{\omega}_1), \cdots, \mathbf{G}(\boldsymbol{\omega}_n))$, $\mathbf{\Omega}_1 = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_n)$, $\mathbf{\Omega}_2 = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n)$, and $\mathbf{\Omega}_{1k}^T$ is the k-th row of $\mathbf{\Omega}_1$.

Missing Data

· Let $\mathbf{V}_{obs} = \{\mathbf{X}_{obs}, \mathbf{Y}_{obs}\}$ and $\mathbf{V}_{mis} = \{\mathbf{X}_{mis}, \mathbf{Y}_{mis}\}$. With $\mathbf{Y} = (\mathbf{Y}_{mis}, \mathbf{Y}_{obs})$ and $\mathbf{V} = (\mathbf{V}_{mis}, \mathbf{V}_{obs})$ given, the conditional distributions corresponding to $\boldsymbol{\theta}$ and $\mathbf{\Omega}$ can be derived in the same way with fully observed data. We only need to derive the conditional distribution corresponding to $\mathbf{V}_{\text{mis}}$.

· (Non-ignorable missing) We define a missing indicator $\mathbf{r}_i = (r_{i1}, \cdots, r_{ip})^T$. If the distribution of $\mathbf{r}$ is independent of $\mathbf{V}_{mis}$, the missing mechanism is defined to be MAR; otherwise the missing mechanism is nonignorable. $p(\mathbf{r}_i|\mathbf{V}, \mathbf{\Omega}, \boldsymbol{\varphi}) = \prod_{i=1}^n\prod_{j=1}^p\{\text{pr}(r_{ij} = 1|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\}^{r_{ij}}\{1 - \text{pr}(r_{ij} = 1|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\}^{1 - r_{ij}}$ where $\boldsymbol{\varphi}$ is parameters in missing data model. Consider logistic model $\text{logit}\{\text{pr}(\mathbf{r_{ij}} = \mathbf{1}|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 v_{i1} + \cdots + \varphi_p v_{ip} + \varphi_{p+1}\omega_{i1} + \cdots + \varphi_{p+q}\omega_{iq} = \boldsymbol{\varphi}^T\mathbf{e}_i$. In the posterior analysis, we iteratively sample from $p(\mathbf{\Omega}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r})$, $p(\mathbf{V}_{mis}|\mathbf{V}_{obs}, \mathbf{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r})$, $p(\boldsymbol{\varphi}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \mathbf{\Omega}, \boldsymbol{\theta}, \mathbf{r})$, $p(\boldsymbol{\theta}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \mathbf{\Omega}, \boldsymbol{\varphi}, \mathbf{r})$.

Two-level SEM

· Consider measurement equation $\mathbf{u}_{gi} = \mathbf{v}_g + \mathbf{\Lambda}_{1g}\boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}$, $g = 1, \cdots, G$, $i = 1, \cdots, N_g$, $\mathbf{v}_g = \boldsymbol{\mu} + \mathbf{\Lambda}_2\boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}$, $g = 1, \cdots, G$. Note that $\mathbf{u}_{gi}$ and $\mathbf{u}_{gj}$ are not independent due to the existence of $\mathbf{v}_g$. And consider structural equation $\boldsymbol{\eta}_{1gi} = \mathbf{\Pi}_{1g}\boldsymbol{\eta}_{1gi} + \mathbf{\Gamma}_{1g}\mathbf{F}_1(\boldsymbol{\xi}_{1gi}) + \boldsymbol{\delta}_{1gi}$, and $\boldsymbol{\eta}_{2g} = \mathbf{\Pi}_2\boldsymbol{\eta}_{2g} + \mathbf{\Gamma}_2\mathbf{F}_2(\boldsymbol{\xi}_{2g}) + \boldsymbol{\delta}_{2g}$. With the Gibbs sampler, we iteratively sample from the following conditional distributions : $[\mathbf{V}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{\Omega}_1, \mathbf{\Omega}_2, \mathbf{X}, \mathbf{Z}]$, $[\mathbf{\Omega}_1|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \mathbf{\Omega}_2, \mathbf{X}, \mathbf{Z}]$, $[\mathbf{\Omega}_2|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \mathbf{\Omega}_1, \mathbf{X}, \mathbf{Z}]$, $[\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}, \mathbf{V}, \mathbf{\Omega}_1, \mathbf{\Omega}_2, \mathbf{X}, \mathbf{Z}]$, $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Y}, \mathbf{V}, \mathbf{\Omega}_1, \mathbf{\Omega}_2, \mathbf{X}, \mathbf{Z}]$.

Multisample Data

· Let $\mathbf{v}_i^{(g)}$ be the $p \times 1$ random vector of observed variables that correspond to the i-th observation (subject) in the g-th group. Consider $\mathbf{v}_i^{(g)} = \boldsymbol{\mu}^{(g)} + \mathbf{\Lambda}^{(g)}\boldsymbol{\omega}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)}$, $\boldsymbol{\eta}_i^{(g)} = \mathbf{\Pi}^{(g)}\boldsymbol{\eta}_i^{(g)} + \mathbf{\Gamma}^{(g)}\mathbf{F}(\boldsymbol{\xi}_i^{(g)}) + \boldsymbol{\delta}_i^{(g)}$. In contrast to two-level SEMs, for $i = 1, \cdots, N_g$ in the g th group, $\mathbf{v}_i^{(g)}$ are assumed to be independent.

· **Identifiability :** When handling ordered categorical outcomes, we impose restrictions on the threshoulds as before. To let underlying latent continuous variables have the same scale among the groups, we select the first group as the reference group and let $\alpha_{m,k}^{(g)} = \alpha_{m,k}^{(1)}$, $k = 1, \cdots, b_m$ for any m.

· (Testing Invariance by model comparison) For unconstrained parameters, we need to specify their own prior distribution, and the data in the corresponding group are used. For constrained parameters across groups, only one prior distribution is needed, and all the data should be used.

Mixture SEM

· Let $\mathbf{y}_i$ be a $p \times 1$ random vector corresponding to the i th observation, and its distribution is $f(\mathbf{y}_i|\boldsymbol{\theta}) = \sum_{k=1}^K \pi_k f_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\theta}_k)$, $i = 1, \cdots, n$ where K is a given integer, π_k is the unknown mixing proportion such that $\pi_k > 0$ and $\pi_1 + \cdots + \pi_K = 1$, $f_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\theta}_k)$ is the multivariate normal density function. For the k-th component, the measurement equation of the model is given by $\mathbf{y}_i = \boldsymbol{\mu}_k + \mathbf{\Lambda}_k\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$, $\boldsymbol{\eta}_i = \mathbf{\Pi}_k\boldsymbol{\eta}_i + \mathbf{\Gamma}_k\boldsymbol{\xi}_i + \boldsymbol{\delta}_i$.

· **Identifiability :** (1) If $\mu_{1,1} < \cdots < \mu_{K,1}$ are well separated, we impose the ordering $\mu_{1,1} < \cdots < \mu_{K,1}$ for solving the label switching problem. (2) If $\mu_{1,1} < \cdots < \mu_{K,1}$ are close to each other, we use random permutation sampler. (3) For each $k = 1, \cdots K$, the SEM is identified by fixing appropriate elements in $\mathbf{\Lambda}_k, \mathbf{\Pi}_k$, and/or $\mathbf{\Gamma}_k$ at preassigned values.

· (Random Permutation) Let $\boldsymbol{\psi} = (\Omega, \mathbf{W}, \boldsymbol{\theta})$, the permutation sampler for generating $\boldsymbol{\psi}$ from the posterior $p(\boldsymbol{\psi}|\mathbf{Y})$ is implemented as follows : (1) Generate $\tilde{\boldsymbol{\psi}}$ from the unconstrained posterior $p(\boldsymbol{\psi}|\mathbf{Y})$ using standard Gibbs sampling steps ; (2) Select some permutation $\rho(1), \cdots, \rho(K)$ and define $\boldsymbol{\psi} = \rho(\tilde{\boldsymbol{\psi}})$ from $\tilde{\boldsymbol{\psi}}$ by reordering the labeling through this permutation : $(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_K) := (\boldsymbol{\theta}_{\rho(1)}, \cdots, \boldsymbol{\theta}_{\rho(K)})$, and $\mathbf{W} = (w_1, \cdots, w_n) := (\rho(w_1), \cdots, \rho(w_n))$.

· **Bayes Analysis :** We introduce a group label w_i for the i-th observation $\mathbf{y}_i$ as a latent allocation variable, and assume $p(w_i = k) = \pi_k$, for $k = 1, \cdots, K$. Let $\boldsymbol{\theta}_{yk}$ be the unknown parameters in $\mathbf{\Lambda}_k$ and $\mathbf{\Psi}_k$, $\boldsymbol{\theta}_{\omega k}$ be the unknown parameters in $\mathbf{\Pi}_k, \mathbf{\Gamma}_k, \mathbf{\Phi}_k$, and $\mathbf{\Psi}_{\delta k}$, $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\pi}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\omega)$. The Gibbs sampler for simulating observations from $[\boldsymbol{\theta}, \mathbf{\Omega}, \mathbf{W}|\mathbf{Y}]$ is : at the r-th iteration with current values $\boldsymbol{\theta}^{(r)}, \mathbf{\Omega}^{(r)}$, and $\mathbf{W}^{(r)}$: Generate $(\mathbf{W}^{(r+1)}, \mathbf{\Omega}^{(r+1)})$ from $p(\mathbf{\Omega}, \mathbf{W}|\mathbf{Y}, \boldsymbol{\theta}^{(r)})$; Generate $\boldsymbol{\theta}^{(r+1)}$ from $p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{\Omega}^{(r+1)}, \mathbf{W}^{(r+1)})$; Finally reorder the label through the permutation sampler to achieve the identifiability.

· **Conditional Distributions :** $p(\mathbf{W}|\mathbf{Y}, \boldsymbol{\theta}) = \prod_{i=1}^n p(w_i|\mathbf{y}_i, \boldsymbol{\theta}) = \prod_{i=1}^n \frac{\pi_k f_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\theta}_k)}{f(\mathbf{y}_i|\boldsymbol{\theta})}$. Let $\mathbf{C}_k = \mathbf{\Sigma}_{\omega k}^{-1} + \mathbf{\Lambda}_k^T\mathbf{\Psi}_k^{-1}\mathbf{\Lambda}_k$, where $\mathbf{\Sigma}_{\omega k}$ is the covariance matrix of $\boldsymbol{\omega}_i$ in the k-th component. $[\boldsymbol{\omega}_i|\mathbf{y}_i, w_i = k, \boldsymbol{\theta}] \stackrel{D}{=} N[\mathbf{C}_k^{-1}\Lambda_k^T\mathbf{\Psi}_k^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_k), \mathbf{C}_k^{-1}]$.

$p(\boldsymbol{\theta}|\mathbf{W}, \mathbf{\Omega}, \mathbf{Y}) = p(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\omega|\mathbf{W}, \mathbf{\Omega}, \mathbf{Y})$
$\propto p(\boldsymbol{\pi})p(\boldsymbol{\mu})p(\boldsymbol{\theta}_y)p(\boldsymbol{\theta}_\omega)p(\mathbf{W}, \mathbf{\Omega}, \mathbf{Y}|\boldsymbol{\theta})$
$\propto p(\boldsymbol{\pi})p(\boldsymbol{\mu})p(\boldsymbol{\theta}_y)p(\boldsymbol{\theta}_\omega)p(\mathbf{W}|\boldsymbol{\theta})p(\mathbf{\Omega}, \mathbf{Y}|\boldsymbol{\theta}, \mathbf{W})$
$\propto p(\boldsymbol{\pi})p(\boldsymbol{\mu})p(\boldsymbol{\theta}_y)p(\boldsymbol{\theta}_\omega)p(\mathbf{W}|\boldsymbol{\pi})p(\mathbf{\Omega}|\mathbf{W}, \boldsymbol{\theta}_\omega)p(\mathbf{Y}|\mathbf{W}, \mathbf{\Omega}, \boldsymbol{\mu}, \boldsymbol{\theta}_y)$
$= [p(\boldsymbol{\pi})p(\mathbf{W}|\boldsymbol{\pi})][p(\boldsymbol{\mu})p(\boldsymbol{\theta}_y)p(\mathbf{Y}|\mathbf{W}, \mathbf{\Omega}, \boldsymbol{\mu}, \boldsymbol{\theta}_y)][p(\boldsymbol{\theta}_\omega)p(\mathbf{\Omega}|\mathbf{W}, \boldsymbol{\theta}_\omega)]$.

$p(\boldsymbol{\pi}) = \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K}\pi_1^\alpha \cdots \pi_K^\alpha$, $p(\boldsymbol{\pi}|\cdot) \propto p(\boldsymbol{\pi})p(\mathbf{W}|\boldsymbol{\pi}) \propto \prod_{k=1}^K \pi_k^{n_k + \alpha}$. Let $\mathbf{Y}_k$ and $\mathbf{\Omega}_k$ be the respective submatrices of $\mathbf{Y}$ and $\mathbf{\Omega}$, such that all the i th columns with $w_i \neq k$ are deleted we have $p(\boldsymbol{\mu}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_\omega|\mathbf{Y}, \mathbf{\Omega}, \mathbf{W}) \propto \prod_{k=1}^K p(\boldsymbol{\mu}_k)p(\boldsymbol{\theta}_{yk})p(\boldsymbol{\theta}_{\omega k})p(\mathbf{Y}_k|\mathbf{\Omega}_k, \boldsymbol{\mu}_k, \boldsymbol{\theta}_{yk})p(\mathbf{\Omega}_k|\boldsymbol{\theta}_{\omega k})$.

Modified Mixture SEM

· A modified DIC : $\text{DIC} = -4E_{\boldsymbol{\theta}_*, \mathbf{F}_m}\{\log p(\mathbf{F}_o, \mathbf{F}_m|\boldsymbol{\theta}_*)|\mathbf{F}_o\} + 2E_{\mathbf{F}_m}\{\log p(\mathbf{F}_o, \mathbf{F}_m|E_{\boldsymbol{\theta}_*}[\boldsymbol{\theta}_*|\mathbf{F}_o, \mathbf{F}_m])|\mathbf{F}_o\}$.

· $\log p(\mathbf{y}_i, \boldsymbol{\omega}_i, \mathbf{d}_i, z_i, \mathbf{x}_i, \mathbf{r}_i^y, \mathbf{r}_i^d, \mathbf{r}_i^x|\boldsymbol{\theta}_*) = \log(\mathbf{y}_i|\boldsymbol{\omega}_i, \boldsymbol{\mu}_k, \mathbf{\Lambda}_k, \mathbf{\Psi}_k, z_i = k) + \log p(\boldsymbol{\eta}_i|\boldsymbol{\xi}_i, \mathbf{d}_i, \mathbf{\Lambda}_{\omega k}, \mathbf{\Psi}_{\delta k}, z_i = k) + \log p(\boldsymbol{\xi}_i|\mathbf{\Phi}_k, z_i = k) + \log p(\mathbf{d}_i|\boldsymbol{\tau}_{kd}, z_i = k) + \log p(z_i = k|\boldsymbol{\tau}, \mathbf{x}_i) + \log p(\mathbf{x}_i|\boldsymbol{\tau}_x) + \log p(\mathbf{r}_i^y|\mathbf{y}_i, \boldsymbol{\varphi}_{ky}, z_i = k) + \log p(\mathbf{r}_i^d|\mathbf{d}_i, \boldsymbol{\varphi}_{kd}, z_i = k) + \log p(\mathbf{r}_i^x|\boldsymbol{\varphi}_x, \mathbf{x}_i) = -\frac{1}{2}\{p\log(2\pi) + \log|\mathbf{\Psi}_k| + (\mathbf{y}_i - \boldsymbol{\mu}_k - \mathbf{\Lambda}_k\boldsymbol{\omega}_i)^T\mathbf{\Psi}_k^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_k - \mathbf{\Lambda}_k\boldsymbol{\omega}_i)\} - \frac{1}{2}\{q_1\log(2\pi) + \log|\mathbf{\Psi}_{\delta k}| + (\boldsymbol{\eta}_i - \mathbf{\Lambda}_{\omega k}\mathbf{G}(\boldsymbol{\omega}_i))^T\mathbf{\Psi}_{\delta k}^{-1}(\boldsymbol{\eta}_i - \mathbf{\Lambda}_{\omega k}\mathbf{G}(\boldsymbol{\omega}_i))\} - \frac{1}{2}\{q_2\log(2\pi) + \log|\mathbf{\Phi}_k| + \boldsymbol{\xi}_i^T\mathbf{\Phi}_k^{-1}\boldsymbol{\xi}_i\} + \log p(\mathbf{d}_i|\boldsymbol{\tau}_{kd}, z_i = k) + \boldsymbol{\tau}_k^T\mathbf{x}_i - \log\{\sum_{j=1}^K \exp(\boldsymbol{\tau}_j^T\mathbf{x}_i)\} + \log p(\mathbf{x}_i|\boldsymbol{\tau}_x) + (\sum_{j=1}^p r_{ij}^y)(\boldsymbol{\varphi}_{ky}^T\mathbf{u}_i^y) - p\log\{1 + \exp(\boldsymbol{\varphi}_{ky}^T\mathbf{u}_i^y)\} + (\sum_{j=1}^{m_2} r_{ij}^d)(\boldsymbol{\varphi}_{kd}^T\mathbf{u}_i^d) - m_2\log\{1 + \exp(\boldsymbol{\varphi}_{kd}^T\mathbf{u}_i^d)\} + (\sum_{j=1}^{m_1} r_{ij}^x)(\boldsymbol{\varphi}_x^T\mathbf{u}_i^x) - m_1\log\{1 + \exp(\boldsymbol{\varphi}_x^T\mathbf{u}_i^x)\}$

$\varphi_x \to r_x$ $\quad \varphi_y \to r_y$ $\quad \tau_x \to X \to Z \to \Omega$, $\Omega \to Y$, $\tau \to$, $\to d \leftarrow \tau_d$

$u_i^x = (1, X_i^T)^T$

$D = \{Y, X, U\}$ $\quad Z$: continuous latent variables $\quad \Omega$: longitudinal latent variables

θ, (u, s) $\quad \{Z, \theta, \Omega, D\}$. $\quad p(\theta, \alpha, Z, \theta, \Omega|D)$