Motivation of SEM:

① Group highly correlated observed variables into latents
② access interrelationships among latents through a regression

1 choose prior

1. high conf: $\Lambda_{0k} \approx \Lambda_k$, small $H_{0yk} = 0.5I$
   otherwise $\Lambda_{0k} = 0$, large $H_{0yk} = 10I$
2. $\Lambda_k^T \omega_i$ is good predictor of y_{ik}, $\psi_{\epsilon k}$ has small E and Var ↓
3. High conf: $\Phi_0 \approx \Phi \Rightarrow R_0^{-1} = (\rho_0 - q_2 - 1)\Phi_0$
4. $\Lambda_{0k}, \Lambda_{0\omega k}, \Phi$ important that link estimates of $\Lambda_k\ \Lambda_{\omega k}, \Phi$ tgt
5. Source: ① expert ② Small data: noninformative
   $p(\Lambda, \psi_\epsilon) \propto \prod_{k=1}^{p} \psi_{\epsilon k}^{-1}$, $p(\Lambda_\omega, \psi_\delta) \propto \prod_{k=1}^{q_1} \psi_{\delta k}^{-1}$, $p(\Phi) \propto |\Phi|^{-\frac{(q_2+1)}{2}}$
   ③ large data: D_0 + noninfo → $p(\theta|D_0)$ as prior
   $D \backslash D_0 + p(\theta|D_0) \to p(\theta|D)$
   ④ moderate data: D + noninfo → $p(\theta|D)$ as prior
   $D + p(\theta|D) \to p(\theta|D)$

$\theta \sim Gamma(\alpha,\beta)$ $p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$ $\mu = \alpha/\beta$ $\sigma^2 = \alpha/\beta^2$

$\theta \sim IG(\alpha,\beta)$ $p(\theta) = \cdots\ \theta^{-(\alpha+1)}e^{-\beta/\theta}$ $\mu = \beta/(\alpha-1)$ $\sigma^2 = \beta^2/(\alpha-1)^2(\alpha-2)$

$W \sim W_q(R_0, \rho_0)$ $p(W) \propto |W|^{(\rho_0-q-1)/2} \times \exp\{-\frac{1}{2}tr(R_0^{-1}W)\}$

$W \sim IW_q(R_0^{-1}, \rho_0)$ $p(W) \propto |W|^{-(\rho_0-q-1)/2} \times \exp\{-\frac{1}{2}tr(R_0^{-1}W^{-1})\}$

$EW = \rho_0 R_0$ $EW^{-1} = R_0^{-1}/(\rho_0 - q - 1)$

2. Derive posterior $\begin{cases} y_i = \Lambda\omega_i + \epsilon_i\ ;\ v_i = (d_i^T, \eta_i^T, \xi_i^T)^T_{r_2+q_1+q_2} \\ \eta_i = Bd_i + \Pi\eta_i + \Gamma\xi_i + \delta_i = \Lambda_\omega v_i + \delta_i \end{cases}$

■ $p(\Omega|Y,\theta) = \prod_{i=1}^n p(\omega_i|y_i,\theta) \propto \prod_{i=1}^n p(\omega_i|\theta)p(y_i|\omega_i,\theta)$

$[\omega_i|\theta] \sim N(\mu_\omega, \Sigma_\omega)$

$\mu_\omega = \begin{bmatrix}\Pi_0^{-1}Bd_i \\ 0\end{bmatrix}$, $\Sigma_\omega = \begin{bmatrix}\Pi_0^{-1}(\Gamma\Phi\Gamma^T + \psi_\delta)\Pi_0^{-T} & \Pi_0^{-1}\Gamma\Phi \\ \Phi\Gamma^T\Pi_0^{-T} & \Phi\end{bmatrix}$

$[y_i|\omega_i,\theta] \sim N(\Lambda\omega_i, \psi_\epsilon)$

$[\omega_i|y_i,\theta] \sim N(\Sigma^{*-1}(\Lambda^T\psi_\epsilon^{-1}y_i + \Sigma_\omega^{-1}\mu_\omega), \Sigma^{*-1})$

$\Sigma^* = \Sigma_\omega^{-1} + \Lambda^T\psi_\epsilon^{-1}\Lambda$

■ $p(\theta|Y,\Omega) \propto p(\theta)p(Y,\Omega|\theta)$ Assume $p(\theta) = p(\theta_y)p(\theta_\omega)$

$p(Y|\Omega,\theta) = p(Y|\Omega,\theta_y)$, $p(\Omega|\theta) = p(\Omega|\theta_\omega)$

• $p(\theta_y|Y,\Omega) \propto p(Y|\Omega,\theta_y)p(\theta_y) = p(\psi_\epsilon)p(\Lambda|\psi_\epsilon)p(Y|\theta_y,\Omega)$

$[\psi_{\epsilon k}^{-1}] \sim Gamma(\alpha_{0\epsilon k}, \beta_{0\epsilon k})$ $[\Lambda_k|\psi_{\epsilon k}] \sim N(\Lambda_{0k}, \psi_{\epsilon k}H_{0yk})$

$[y_i|\theta_y,\omega_i] \sim N(\Lambda\omega_i, \psi_\epsilon)$

$[\psi_{\epsilon k}^{-1}|Y,\Omega] \sim Gamma(\frac{n}{2} + \alpha_{0\epsilon k}, \beta_{\epsilon k})$

$[\Lambda_k|Y,\Omega,\psi_{\epsilon k}^{-1}] \sim N(a_k, \psi_{\epsilon k}A_k)$

$\beta_{\epsilon k} = \beta_{0\epsilon k} + \frac{1}{2}(Y_k^TY_k - a_k^TA_k^{-1}a_k + \Lambda_{0k}^TH_{0yk}^{-1}\Lambda_{0k})$

$a_k = A_k(H_{0yk}^{-1}\Lambda_{0k} + \Omega Y_k)$; $A_k = (H_{0yk}^{-1} + \Omega\Omega^T)^{-1}$

If fixed $L = [l_{kj}]$, $l_{kj} = 0$ fixed or 1 free

Λ_k^* vector of unknown params in Λ_k

Y_k submatrix of Y delete row with $l_{kj} = 0$.

$Y_k^{*T} = (y_{1k}^*, \cdots, y_{nk}^*)$, $y_{nk}^* = y_{ik} - \sum_{j=1}^q \lambda_{kj}\omega_{ij}(1-l_{kj})$

Replace Λ_k & Y_k with Λ_k^* and Y_k^*, respectively.

• $p(\theta_\omega|Y,\Omega) \propto p(\Omega|\theta_\omega)p(\theta_\omega) = p(\Omega_1|\Omega_2,\theta_\omega)p(\Omega_2|\theta_\omega)p(\theta_\omega)$

$= [p(\Omega_1|\Omega_2, B, \Pi, \Gamma, \psi_\delta)p(B,\Pi,\Gamma,\psi_\delta)][p(\Omega_2|\Phi)p(\Phi)]$

$[\Phi] \sim IW_{q_2}(R_0^{-1}, \rho_0)$ $[\Omega_2|\Phi] \sim N(0, \Phi)$

$[\Phi|\Omega_2] \sim IW_{q_2}((\Omega_2\Omega_2^T + R_0^{-1}), n + \rho_0)$

$\Lambda_\omega = [B, \Pi, \Gamma]$, $v_i = (d_i^T, \eta_i^T, \xi_i^T)^T$ similar to $y_i = \Lambda\omega_i + \epsilon_i$

Basic Assumption $\epsilon \perp\!\!\!\perp \omega$

$\epsilon_i \overset{iid}{\sim} N(0, \psi_\epsilon)$, $\psi_\epsilon = diag\{\psi_{\epsilon k}\}_1^p$

$\delta_i \overset{iid}{\sim} N(0, \psi_\delta)$, $\psi_\delta = diag\{\psi_{\delta k}\}_1^{q_1}$

$\xi_i \overset{iid}{\sim} N(0, \Phi)$, Φ general $q_2 \times q_2$

$\mu \sim N(\mu_0, \Sigma_0)$

$\psi_{\epsilon k} \sim IG(\alpha_{0\epsilon k}, \beta_{0\epsilon k})$

$\Lambda_k|\psi_{\epsilon k} \sim N(\Lambda_{0k}, \psi_{\epsilon k}H_{0yk})$

$\Phi \sim IW_{q_2}(R_0^{-1}, \rho_0)$

$\psi_{\delta k} \sim IG$, $\Lambda_{\omega k}|\psi_{\delta k} \sim N$

3. sample for posterior (Metropolis-Hastings algorithm)

(1) Data augmentation: $p(\theta|Y) \to p(\theta,\Omega|Y)$

(2) Gibbs sampling: ① initiate $\theta^{(0)}$ & $\Omega^{(0)}$

② $\theta^{(j+1)} \leftarrow p(\theta|\Omega^{(j)}, Y)$

$\Omega^{(j+1)} \leftarrow p(\Omega|\theta^{(j+1)}, Y)$

(3) Check convergence ① Plot convergence curves from distinct initials

② Estimated potential scale reduction (EPSR) value < 1.2

$B = \frac{n}{k-1}\sum_{k=1}^K(\bar\theta_{\cdot k} - \bar\theta_{\cdot\cdot})^2$ $\bar\theta_{\cdot k} = \frac{1}{n}\sum\theta_{jk}$ $\bar\theta_{\cdot\cdot} = \frac{1}{k}\sum\bar\theta_{\cdot k}$

$W = \frac{1}{K}\sum_{k=1}^K S_k^2$

$S_k^2 = \frac{1}{n-1}\sum_{j=1}^n(\theta_{jk} - \bar\theta_{\cdot k})^2$ $\widehat{Var}\theta = \frac{n-1}{n}W + \frac{1}{n}B$ $\hat R^{\frac{1}{2}} = [\widehat{Var}\theta/W]^{\frac{1}{2}}$

(4) Statistical Inference

$\hat\theta = \frac{1}{T}\sum_{t=1}^T\theta^{(t)}$ $\widehat{Var}(\theta|Y) = \frac{1}{T-1}\sum_{t=1}^T(\theta^{(t)} - \hat\theta)(\theta^{(t)} - \hat\theta)^T$

$\hat\omega_i = \frac{1}{T}\sum_{t=1}^T\omega_i^{(t)}$

(5) Sensitivity analysis: priors for $\Lambda, \Lambda_\omega, \psi_\epsilon, \psi_\delta, \Phi$

① different priors ② ad hoc priors

Model comparison

1. Bayes Factor $B_{10} = \frac{p(Y|M_1)}{p(Y|M_0)} = \frac{p(Y|M_1)/p(Y|M_a)}{p(Y|M_0)/p(Y|M_a)}$
   $\cdots = B_{1a}/B_{0a}$

① $p(\Omega,\theta|Y,t) = \frac{1}{z(t)}p(Y,\Omega,\theta|t)$, $z(t) = p(Y|t)$

② $\frac{d\log z(t)}{dt} = \int\frac{1}{z(t)}\cdot\frac{d}{dt}p(Y,\Omega,\theta|t)d\Omega d\theta$
$= E_{\Omega,\theta}[\frac{d}{dt}\log p(Y,\Omega,\theta|t)]$

$U(Y,\Omega,\theta|t) = \frac{d}{dt}\log p(Y,\Omega|\theta,t)$

$\Rightarrow \log B_{10} = \log\frac{z(1)}{z(0)} = \int_0^1\frac{d}{dt}\log z(t)dt = \int_0^1 E_{\Omega,\theta}[U(\cdots|t)]dt$

③ $\widehat{\log B_{10}} = \frac{1}{2}\sum_{s=0}^S(t_{(s+1)} - t_{(s)})(\bar U_{(s+1)} + \bar U_{(s)})$

$\bar U_{(s)} = \frac{1}{J}\sum_{j=1}^J U(Y, \Omega^{(j)}, \theta^{(j)}, t_{(s)})$ $S=20, J=1000$

Example

M_t: $\eta_i = bd_i + \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + (1-t)\gamma_{12}\xi_{i1}\xi_{i2} + t\gamma_{22}\xi_{i2}^2 + \delta_i$

For each $t = t_{(s)}$, $\{(\Omega^{(j)}, \theta^{(j)}), j=1,\cdots,J\} \leftarrow N$.

$U(Y, \Omega^{(j)}, \theta^{(j)}, t_{(s)}) = \frac{d}{dt}(\log p(Y,\Omega,\theta,t))\Big|_{t=t_{(s)}}$

$= \sum_{i=1}^n\underbrace{\{\eta_i - \cdots - t\gamma_{22}\xi_{i2}^2\}}_{\text{residual}}\underbrace{(-\gamma_{12}\xi_{i1}\xi_{i2} + \gamma_{22}\xi_{i2}^2)}_{\frac{d}{dt}\cdots}\underbrace{/\psi_\delta}_{\Sigma}$

Calculate $\bar U_{(s)}$, $s = 0,\cdots,S+1$, $\Rightarrow \widehat{\log B_{10}}$

2. DIC (Deviance Information Criterion) smaller better

$DIC = -4E_{\theta k}[\log p(Y|\theta_k, M_k)|Y] + 2\log p(Y|\bar\theta_k)$ Bayesian E.

$\approx \frac{1}{J}\sum_{j=1}^J\log(Y|\theta_k^{(j)}, M_k)$

* mixture model does not give DIC values in WinBUGS

* $\Delta DIC < 5$, then we should not report who is better.

$[\psi_{\delta k}^{-1}|\Omega] \sim Gamma(\frac{n}{2} + \alpha_{0\delta k}, \beta_{\delta k})$ $E_{ik}^T = (\eta_{1k}^*, \cdots, \eta_{nk}^*)$

$[\Lambda_{\omega k}|\Omega, \psi_{\delta k}^{-1}] \sim N(a_{\omega k}, \psi_{\delta k}A_{\omega k})$

$\beta_{\delta k} = \beta_{0\delta k} + \frac{1}{2}(E_{ik}^TE_{ik} - a_{\omega k}^TA_{\omega k}^{-1}a_{\omega k} + \Lambda_{0\omega k}^TH_{0\omega k}^{-1}\Lambda_{0\omega k})$

$a_{\omega k} = A_{\omega k}(H_{0\omega k}^{-1}\Lambda_{0\omega k} + V_kE_{ik})$; $A_{\omega k} = (H_{0\omega k}^{-1} + V_kV_k^T)^{-1}$

3 L_ν-measure 1. means of posterior predictive distribution 2. variances

$L_\nu(Y) = \sum_{i=1}^n tr[Cov(y_i^{rep}|Y)] + \nu\sum_{i=1}^n tr\{[E(y_i^{rep}|Y) - y_i][E(y_i^{rep}|Y) - y_i]^T\}$

$Cov(y_i^{rep}|Y) = E[Cov(y_i^{rep}|\Omega,\theta)|Y] + Cov[E(y_i^{rep}|\Omega,\theta)|Y]$

$E[E(y_i^{rep}|\Omega,\theta)|Y]$

$\widehat{Cov}(y_i^{rep}|Y) = \frac{1}{J}\sum_{j=1}^J[\Lambda_\eta^{(j)}(\Pi_0^{(j)})^{-1}\psi_\delta^{(j)}(\Lambda_\eta^{(j)}(\Pi_0^{(j)})^{-1})^T + \psi_e^{(j)}]$
$+ \frac{1}{J}\sum_{j=1}^J m_i^{(j)}m_i^{(j)T} - (\frac{1}{J}\sum_{j=1}^J m_i^{(j)})(\frac{1}{J}\sum_{j=1}^J m_i^{(j)})^T$

$m_i^{(j)} = \mu^{(j)} + \Lambda_\eta^{(j)}(\Pi_{(0)}^{(j)})^{-1}[\Gamma^{(j)}F(\xi_i^{(j)})] + \Lambda_\xi^{(j)}\xi_i^{(j)}$

4 PP p-value (Posterior Predictive p-value).

$p_B(Y) = Pr\{\underbrace{D(Y^{rep}|\theta,\Omega)}_{\sim\chi^2(d^*)} \geq D(Y|\theta,\Omega)|Y, M_0\} \approx 0.5, M_0$

$\Rightarrow p_B(Y) = \int P_r(\chi^2(d^*) \geq D(Y|\theta,\Omega))p(\theta,\Omega|Y)d\theta d\Omega$
$\approx \frac{1}{J}\sum_{j=1}^J Pr\{\chi^2(d^*) \geq D(Y|\theta^{(j)}, \Omega^{(j)})\}$.

5. Residual analysis ME: $\hat\epsilon_i = y_i - \hat A c_i - \hat\Lambda\hat\omega_i$
= or QQ plot. SE: $\hat\delta_i = (I - \hat\Pi)\hat\eta_i - \hat Bd_i - \hat\Gamma\hat\xi_i$

*missing data $v_i = (x_i^T, y_i^T)^T$ latent variables.

\- MAR $\{\Omega, X_{mis}, Y_{mis}, Y_{obs}\} = \{\Omega, V_{mis}, Y_{obs}\}$

Gibbs: $P(\cdot|\overset{①}{\theta}, \overset{④}{\alpha}, \overset{②}{\Omega}, \overset{③}{\boxed{V_{mis}}}, \overset{④}{Y_{obs}}, X_{obs}, Z_{obs}\}$

$p(V_{mis}|\theta,\alpha,\Omega,Y_{obs},X_{obs},Z_{obs}) = \prod_{i=1}^n p(V_{i,mis}|\theta,\omega_i)$

$[v_{i,mis}|\theta,\omega_i] \sim N(\mu_{i,mis} + \Lambda_{i,mis}\omega_i, \psi_{i,mis})$

\- MANR: P(missing) ⫫̸ missing data. $r_{ij} = \begin{cases}1 & v_{ij} \text{ is mis} \\ 0 & v_{ij} \text{ is obs}\end{cases}$

$p(r|V,\Omega,\varphi) = \prod_{i=1}^n\prod_{j=1}^p[P(r_{ij}=1|v_i,\omega_i,\varphi)]^{r_{ij}}[1 - P(\cdots)]^{1-r_{ij}}$

$p(\theta,\varphi|V_{obs},r) \propto \{\prod_{i=1}^n\int_{\omega_i,v_{i,mis}}p(v_i|\omega_i,\theta) \times$
$p(r_i|v_i,\omega_i,\varphi)p(\omega_i|\theta)d\omega_i dv_{i,mis}\}p(\theta,\varphi)$

$m(v_i,\omega_i,\varphi) = logit\{P(r_i|v_i,\omega_i,\varphi)\}$
$= \varphi^Te_i$ $e_i = [1, v_{i1}, \cdots, v_{ip}, \omega_{i1}, \cdots, \omega_{iq}]^T$

Gibbs: $P(\cdot|\Omega, V_{obs}, V_{mis}, \theta, \varphi, r)$

$p(\Omega|\cdot) = P(\Omega|V,\theta,\varphi,r)$; $p(V_{mis}|\cdot)$

$p(\varphi|\cdot) = P(\varphi|V,\Omega,\theta,r)$; $p(\theta|\cdot) = p(\theta|V,\Omega)$

Notation:

$Y = \{y_i, i=1,\cdots,n\}$ $D = \{d_i, i=1,\cdots,n\}$, $X = \{x_i, i=1,\cdots,n\}$

$Z = \{z_i, i=1,\cdots,n\}$, $\Omega = \{\omega_i, i=1,\cdots,n\}$

$y_i = \{y_{oi}, y_{mi}\}$, $d_i = \{d_{oi}, d_{mi}\}$ $x_i = \{x_{oi}, x_{mi}\}$

$Y_o = \{y_{oi}, i=1,\cdots,n\}$, $D_o = \{d_{oi}, i=1,\cdots,n\}$, $X_o = \{x_{oi}, i=1,\cdots,n\}$

$R^y = \{r_i^y, i=1,\cdots,n\}$, $R^d = \{r_i^d, i=1,\cdots,n\}$, $R^x = \{r_i^x, i=1,\cdots,n\}$

$F_o = \{Y_o, D_o, X_o, R^y, R^d, R^x\}$

$Y_m = \{y_{mi}, i=1,\cdots,n\}$, $D_m = \{d_{mi}, i=1,\cdots,n\}$, $X_m = \{x_{mi}, i=1,\cdots,n\}$

$F_m = \{Y_m, D_m, X_m, \Omega, Z\}$

$\pi_k = \{\pi_{ik}, i=1,\cdots,n\}$, $\pi = \{\pi_k, k=1,\cdots,K-1\}$, $\tau = \{\tau_k, k=1,\cdots,K-1\}$

$\tau_d = \{\tau_{kd}, k=1,\cdots,K\}$

$\varphi_y = \{\varphi_{ky}, k=1,\cdots,K\}$, $\varphi_d = \{\varphi_{kd}, k=1,\cdots,K\}$, $\varphi_x = \{\varphi_{kx}, k=1,\cdots,K\}$

$\Rightarrow \begin{cases}\vartheta_s = \{\varphi_y, \varphi_d, \varphi_x, \tau, \tau_x, \tau_d\} \\ \theta_* = \{\theta, \pi, \vartheta_s\}\end{cases}$

*Ordered categorical variables. $\Omega \to X, Y \xrightarrow{\alpha} \boxed{Z}$

Assumption $y \sim N \to z \sim$ OrderCat

by $z = m$ if $\alpha_m \leq y < \alpha_{m+1}$ $m = 0,\cdots,b$

$-\infty = \alpha_{k,0} < \alpha_{k,1} < \cdots < \alpha_{k,b_k} < \alpha_{k,b_k+1} = +\infty$

Identification conditions

① fix some elements in Λ, Π, and Γ

② indeterminacy induced by the orderCat variable.
by imposing $\alpha_1 = \Phi^{-1}(f_1)$, $\alpha_b = \Phi^{-1}(1 - f_b)$
implicitly unifying $y \sim N(0,1)$

*Dichotomous variables: only one threshold.

Assumption $y_{ik} \sim N \to z_{ik} = \begin{cases}1 & \text{if } y_{ik} > 0 \\ 0 & \text{otherwise}\end{cases}$

$P(z_{ik} = 1|\omega_i, \mu_k, \Lambda_k, \psi_{\epsilon k})$
$= \Phi([\mu_k + \Lambda_k^T\omega_i]/\psi_{\epsilon k}^{1/2})$
no threshold needs to be estimated in Gibbs sampler.

Identification conditions ① fix ⋯
② fix $\psi_{\epsilon k} = 1$ for the covariance

Gibbs sampler

(1) $\Omega^{(j+1)} \leftarrow P(\Omega|\theta^{(j)}, \alpha^{(j)}, Y^{(j)}, X, Z)$
(2) $\theta^{(j+1)} \leftarrow p(\theta|\Omega^{(j+1)}, \alpha^{(j)}, Y^{(j)}, X, Z)$
(3) OrdCat $\alpha^{(j+1)}, Y^{(j+1)} \leftarrow P(\alpha, Y|\theta^{(j+1)}, \Omega^{(j+1)}, X, Z)$
Dichot $[y_{ik}|\theta, \omega_i, z_i] \sim$
$\begin{cases}N(\mu_k + \Lambda_k^T\omega_i, 1)\mathbb{1}_{(-\infty,0]}(y_{ik}), & z_{ik} = 0 \\ N(\mu_k + \Lambda_k^T\omega_i, 1)\mathbb{1}_{(0,+\infty)}(y_{ik}), & z_{ik} = 1\end{cases}$

* Exponential family. Assumption

$p(y_{ik}|\omega_i) = \exp\{[y_{ik}\vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\epsilon k} + c_k(y_{ik}, \psi_{\epsilon k})\}$

$E(y_{ik}|\omega_i) = \dot b(\vartheta_{ik})$ $Var(y_{ik}|\omega_i) = \psi_{\epsilon k}\ddot b(\vartheta_{ik})$

$\vartheta_{ik} = A_k^Tc_{ik} + \Lambda_k^T\omega_i$ $\eta_i = Bd_i + \Pi\eta_i + \Gamma F(\xi_i) + \delta_i$

$M_{y_{ik}|\omega_i}(t) = E_{y_{ik}|\omega_i}e^{y_{ik}t}$
$= \int\exp\{[y_{ik}\vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\epsilon k} + c_k(y_{ik}, \psi_{\epsilon k})\}dy_{ik}$
$= \exp\{-b(\vartheta_{ik})/\psi_{\epsilon k}\}\int\exp\{y_{ik}(\vartheta_{ik} + \psi_{\epsilon k}t)/\psi_{\epsilon k} + c_k(y_{ik}, \psi_{\epsilon k})\}dy_{ik}$
$= \exp\{[b(\vartheta_{ik} + \psi_{\epsilon k}t) - b(\vartheta_{ik})]/\psi_{\epsilon k}\}$
$\Rightarrow C(t) = \psi_{\epsilon k}^{-1}[b(\vartheta_{ik} + \psi_{\epsilon k}t) - b(\vartheta_{ik})]$
$\Rightarrow \mu = \dot b(\vartheta_{ik})$, $\sigma^2 = \psi_{\epsilon k}\ddot b(\vartheta_{ik})$
$\exp\{b(\vartheta_{ik})/\psi_{ik}\} = \int\exp\{y_{ik}\vartheta_{ik}/\psi_{\epsilon k} + c_k(y_{ik}, \psi_{\epsilon k})\}dy_{ik}$.

Identification: same with OrdCat.

Prior: $\alpha \perp\!\!\!\perp \theta$, $\alpha_k \perp\!\!\!\perp \alpha_l$ (Next Page)
$p(\alpha_k) = p(\alpha_{k,2}, \cdots, \alpha_{k,b_k-1}) \propto c$ noninfo.

BF in EFD: $D_o = (X, Z)$. $D_m = (Y, \Omega)$, $\theta^* = (\theta, \alpha)$

$U(\theta^*, D_m, D_o, t) = \frac{d}{dt}\log p(D_m, D_o|\theta^*, t)$

$\bar U$ & $\widehat{\log B_{10}}$ are same with the standard one.

Example $\vec v_i = (\underbrace{z_{i1}, z_{i2}, z_{i3}}_{dich}, \underbrace{x_{i4}, \cdots, x_{iq}}_{Bernoulli})^T$

$x_{ik} \propto \exp\{x_{ik}\vartheta_{ik} - \underbrace{\log(1 + e^{\vartheta_{ik}})}_{b(\vartheta_{ik})}\}$ $\vartheta_{ik} = \log(\frac{p_{ik}}{1 - p_{ik}})$

$y_{ik} = \mu_k + \Lambda_k^T\omega_i + \epsilon_i$ $\vartheta_{ik} = \mu_k + \Lambda_k^T\omega_i$ $(\frac{SE}{SD} > 1)$

① large sample size needed ② $\sqrt{\widehat{Var}(\theta|V)}$ overest true SD

③ incorrectly treat dichotomous as binary or vice versa would produce misleading result.

$Mean(\hat\theta) = \frac{1}{R}\sum\hat\theta_r$, MC-SD$(\hat\theta)$ $\sqrt{\frac{1}{R-1}\sum(\hat\theta_r - \theta)^2}$

EST-SD$(\hat\theta) = \sqrt{\widehat{Var}(\theta|V)}$ RMS$(\hat\theta) = \sqrt{\frac{1}{R}\sum_{r=1}^R(\hat\theta_r - \theta)^2}$

SE/SD$(\hat\theta) = \frac{\text{EST-SD}(\hat\theta)}{\text{MC-SD}(\hat\theta)}$

Hierarchical Data

come from a number of diff groups with known hierarchical structure!

L1 (within group) $u_{gi} = v_g + \Lambda_{1g}\omega_{gi} + \varepsilon_{gi}$ $\quad g = 1,\cdots,G,\ i = 1,\cdots,N_g$

L2 (between group) $v_g = \mu + \Lambda_{2g}\omega_{2g} + \varepsilon_{2g}$ $\quad u_{gi} = \begin{bmatrix} x_{gi} \\ y_{gi} \end{bmatrix} \to z_{gi}$

$\Rightarrow u_{gi} = \mu + \Lambda_2\omega_{2g} + \varepsilon_{2g} + \Lambda_{1g}\omega_{1gi} + \varepsilon_{gi}$

$\eta_{1gi} = \Pi_{1g}\eta_{1gi} + \Gamma_{1g}F_1(\xi_{1gi}) + \delta_{1gi}$

$\eta_{2g} = \Pi_2\eta_{2g} + \Gamma_2 F_2(\xi_{2g}) + \delta_{2g}$

Assumption within group params are invariant / shared. coeff & errors are distinct.

Example M_1: $v_g = \mu + \Lambda_2^1\omega_{2g} + \varepsilon_{2g}$ $\quad \omega_{2g} \sim N(0,\Phi_2)$ FA

M_2: $v_g = \mu + \Lambda_2^2\omega_{2g} + \varepsilon_{2g}$

$\eta_{2g} = \Pi_2^2\eta_{2g} + \Gamma_2^2 F_2(\xi_{2g}) + \delta_{2g}$

M_a: $v_g = \mu + \varepsilon_{2g}$

link model M_{t1a}: $u_{gi} = \mu + t\Lambda_2^1\omega_{2g} + \varepsilon_{2g} + \Lambda_1\omega_{gi} + \varepsilon_{gi}$

$\omega_{2g} \sim N(0,\Phi_2)$ $\quad \eta_{1gi} = \Pi_1\eta_{1gi} + \Gamma_1 F_1(\xi_{1gi}) + \delta_{gi}$

M_{t2a}: $u_{gi} = \mu + t\Lambda_2^2\omega_{2g} + \Lambda_1\omega_{gi} + \varepsilon_{gi}$

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

* Multisample data. small G large N_g $\quad Obs_{gi} \perp\!\!\!\perp\ i = 1,\cdots,N_g$

target investigate the similarity or diff among groups

test: H_0: params invariant among groups.

$v_i^{(g)} = \mu^{(g)} + \Lambda^{(g)}\omega_i^{(g)} + \varepsilon_i^{(g)} \sim N(0,\Psi_\varepsilon^{(g)})$

$\Downarrow \begin{bmatrix} x_i^{(g)} \\ y_i^{(g)} \end{bmatrix} \to z_i^{(g)}$ $\qquad \sim N(0,\Phi^{(g)})$

$\eta_i^{(g)} = \Pi^{(g)}\eta_i^{(g)} + \Gamma^{(g)}F(\xi_i^{(g)}) + \delta_i^{(g)} \sim N(0,\Psi_\delta^{(g)})$

Assumption parameters have the same dimension among group.

identification

① fix $\alpha_{k,1}^{(g)}$ & $\alpha_{k,b_k}^{(g)}$ among groups sharing scale

eg. $\alpha_{k,b}^{(g)} = \alpha_{k,b}^{(1)}$, $\quad k = 1,\cdots,s.\ \ b = 1,\cdots,b_k$

Model comparison ① Non constrained params have its own priors for each groups ② Constrained params shares one prior and all data should be combined in estimation ③ For constraint $\Lambda^{(1)} = \cdots = \Lambda^{(G)} = \Lambda$, but $\Psi_\varepsilon^{(1)} \neq \cdots \neq \Psi_\varepsilon^{(G)}$, we can not use joint prior dist for Γ and ε_i: $\Lambda_\omega^{(1)} = \cdots = \Lambda_\omega^{(G)}$, $\Lambda_{\omega k} \sim N(\Lambda_{0\omega k}, H_{0\omega k})$

no $\psi_{\varepsilon k}$ $\to$ $\quad \Lambda_\omega^{(1)} \neq \cdots \neq \Lambda_\omega^{(G)}$, $\Lambda_{\omega k} \sim N(\Lambda_{0\omega k}^{(g)}, H_{0\omega k}^{(g)})$

$(2\pi)^{-\frac{p}{2}}|\Psi_k|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(y_i - \mu_k - \Lambda_k\omega_i)^T\Psi_k^{-1}(\cdots)\}$

$\times(2\pi)^{-\frac{q_1}{2}}|I_{q_1} - \Pi_k||\Psi_{\delta k}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\eta_i - \Lambda_{\omega k}\omega_i)^T\Psi_{\delta k}^{-1}(\cdots)\}$

$\times(2\pi)^{-\frac{q_2}{2}}|\Phi_k|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\xi_i^T\Phi_k^{-1}\xi_i\}$

$\Lambda_{\omega k} = (\Pi_k, \Gamma_k)$

Diff of mixture, hierarchical, and multisample

	K	N_g	group membership	obs $\perp\!\!\!\perp$?
mix	small	large	unknown	$\perp\!\!\!\perp$
Hier	large	small	known	$\not\perp\!\!\!\perp$ ← share certain common influential factor
Mul	small	large	known	$\perp\!\!\!\perp$

*Mixture model $\quad f(y_i|\theta) = \sum_{k=1}^K \pi_k f_k(y_i|\mu_k,\theta_k)$

– general mixture π_k $\quad$ their own prior.

identification condition: $\mu_{1,1} < \cdots < \mu_{K,1}$ solving label switching

Gibbs sampler ① $(W^{(j+1)}, \Omega^{(j+1)}) \leftarrow p(\Omega, W|Y,\theta^{(j)})$

② $\theta^{(j+1)} \leftarrow p(\theta|Y,\Omega^{(j+1)},W^{(j+1)})$

$\rho: \{1..k\} \to \{1..k\}$ ③ Reorder the label by $\psi = \rho(\tilde{\psi})$

$(\theta_1,\cdots,\theta_K) = (\theta_{\rho(1)},\cdots,\theta_{\rho(k)})$, $W = (w_1,\cdots,w_n) := (\rho(w_1),\cdots,\rho(w_n))$

$d_{kh} = \max_{\ell\in\{k,h\}}\{(\mu_k - \mu_h)^T\Sigma_\ell^{-1}(\mu_k - \mu_h)\}^{\frac{1}{2}} \geq 3.8$

Detail. ① $P(\Omega,W|Y,\theta) = p(W|Y,\theta)p(\Omega|Y,W,\theta)$

ⓐ $W^{(j+1)} \leftarrow p(W|Y,\theta^{(j)})$,

$= \prod_{i=1}^n p(w_i|y_i,\theta)$ $\quad p(w_i = k|y_i,\theta) = \frac{\pi_k f_k(y_i|\mu_k,\theta_k)}{f(y_i|\theta)}$ $\quad \sim N(\mu_k, \Sigma_k(\theta_k))$

$\Sigma_k(\theta_k) = \begin{bmatrix} \Pi_{0k}^{-1}(\Gamma_k\Phi_k\Gamma_k^T + \Psi_{\delta k})\Pi_{0k}^{-T} & * \\ \Phi_k\Gamma_k^T\Pi_{0k}^{-T} & \Phi_k \end{bmatrix}$

ⓑ $\Omega^{(j+1)} \leftarrow p(\Omega|Y,\theta^{(j)},W^{(j+1)})$

$= \prod_{i=1}^n p(y_i|\omega_i,w_i,\mu,\theta_y)p(\omega_i|w_i,\theta_\omega)$ $\quad (w_i = k)$

$\sim N(\mu_k + \Lambda_k\omega_i, \Psi_k)$ $\qquad \tilde{N}(C_k^{-1}\Lambda_k^T\Psi_k^{-1}(y_i - \mu_k), C_k^{-1})$

$C_k = \Sigma_{\omega k}^{-1} + \Lambda_k^T\Psi_\varepsilon^{-1}\Lambda_k$

② $p(\theta|Y,\Omega,W)$ Assume

1. prior(π) $\perp\!\!\!\perp$ prior($\mu,\theta_y,\theta_\omega$)
2. prior(μ) $\perp\!\!\!\perp$ prior(θ_y,θ_ω)
3. prior(θ_y) $\perp\!\!\!\perp$ prior(θ_ω)

$p(\theta) = p(\pi)p(\mu)p(\theta_y)p(\theta_\omega)$

$p(\theta|W,\Omega,Y)$

$\propto \underbrace{[p(\pi)p(W|\pi)]}_{MX}\underbrace{[p(\mu)p(\theta_y)p(Y|W,\Omega,\mu,\theta_y)]}_{ME}\underbrace{[p(\theta_\omega)p(\Omega|W,\theta_\omega)]}_{SE}$

$p(\pi|\cdot)$ $\qquad p(\mu,\theta_y|\cdot)$ $\qquad p(\theta_\omega|\cdot)$

$[\pi] \sim D(\alpha,\cdots,\alpha)$, $[W|\pi] \sim Multi(\pi_1,\cdots,\pi_k)$

$[\pi|\cdot] \sim D(\alpha + n_1,\cdots,\alpha + n_k)$

Assume $k \neq h$ $\Rightarrow (\mu_k,\theta_{yk},\theta_{\omega k}) \perp\!\!\!\perp (\mu_h,\theta_{yh},\theta_{\omega h})$

refer to standard Bayesian inference.

Given W, mixture SEM become multisample SEM.

Model comparison

$M_1: Y|\theta,\pi \sim \sum_{k=1}^K$ $\quad M_0: Y|\theta,\pi \sim \sum_{k=1}^{C} \pi_k f_k(Y|\mu_k,\Sigma_k)$

$M_t: y|\theta,\pi \sim [\pi_1 + (1-t)a_1(\pi_{c+1} + \cdots + \pi_K)]f_1(\cdot|\cdot) + \cdots$

$+ [\pi_c + (1-t)a_c(\pi_{c+1} + \cdots + \pi_K)]f_c(\cdot|\cdot) +$

$\sum_{k=1}^c a_k = 1$ or $a_k = \frac{1}{c}$ $\quad + t\pi_{c+1}f_{c+1}(\cdot|\cdot) + \cdots + t\pi_K f_K(\cdot|\cdot)$

$\log p(Y,\Omega|\theta,t) = \sum_{i=1}^n \{\log\{\sum_{k=1}^c[\pi_k + (1-t)a_k\sum_{h=c+1}^K \pi_h]f_k(y_i,\omega_i|\mu_k,\theta_k) + \sum_{k=c+1}^K t\pi_k f_k(y_i,\omega_i|\mu_k,\theta_k)\}$

$U(Y,\Omega,\theta,t) = \sum_{k=1}^n \frac{-\sum_{h=c+1}^K \pi_h\sum_{k=1}^c a_k f_k(y_i,\omega_i|\mu_k,\theta_k) + \sum_{k=c+1}^K \pi_k f_k(\cdot|\cdot)}{\sum_{k=1}^c[\pi_k + (1-t)a_k\sum_{h=c+1}^K \pi_h]f_k(\cdot|\cdot) + \sum_{k=c+1}^K t\pi_k f_k(\cdot|\cdot)}$

$f_k(y_i,\omega_i|\mu_k,\theta_k) = \cdots$

– r.v. modeled by multinomial logit model.

1. exam effects of covariates on component membership
2. capture the component-specific nonlinear interrelationship among explanatory latents or covariates on outcome latents
3. incorporate unignorable missing response and covariates

$\pi_{ik} = p(z_i = k|x_i) = \frac{\exp\{\tau_k^T x_i\}}{\sum_{k=1}^K \exp\{\tau_k^T x_i\}}$

identification $\tau_K = \vec{0}$

Model $y_i = \mu_k + \Lambda_k\omega_i + \varepsilon_i$ $\quad // \Lambda_{\omega k}G(\omega_i) + \delta_i$

$\eta_i = B_k d_i + \Pi_k\eta_i + \Gamma_k F(\xi_i) + \delta_i$

missing indicator vector $r_i^y \in R^p$ $\ r_i^d \in R^{m_2}$ $\ r_i^x \in R^{m_1}$

$\underbrace{p(r_i^y|y_i,\omega_i), p(r_i^d|d_i,\omega_i)}$, $p(r_i^x|x_i,\omega_i)$ diff in every component.

$p(r_i^y|y_i,\omega_i,\varphi_{ky}) = \prod_{j=1}^p[p(r_{ij}^y = 1|y_i,\omega_i,\varphi_{ky})]^{r_{ij}^y}[1 - p(r_{ij}^y = 1|y_i,\omega_i,\varphi_{ky})]^{1-r_{ij}^y}$

$\text{logit}\{p(r_{ij}^y = 1|y_i,\varphi_{ky})\} = \varphi_{k0}^y + \varphi_{k1}^y y_{i1} + \cdots + \varphi_{kp}^y y_{ip} = \varphi_{ky}^T u_i^y$, $\quad u_i^y = [1, y_i^T]^T$. $\ \varphi_{ky} = (\varphi_{k0}^y, \varphi_{k1}^y, \cdots, \varphi_{kp}^y)^T$

$p(r_i^d|d_i,\omega_i,\varphi_{kd}) = \prod_{j=1}^{m_2}[p(r_{ij}^d = 1|d_i,\omega_i,\varphi_{kd})]^{r_{ij}^d}[1 - p(r_{ij}^d = 1|d_i,\omega_i,\varphi_{kd})]^{1-r_{ij}^d}$

$\text{logit}\{p(r_{ij}^d = 1|d_i,\varphi_{kd})\} = \cdots = \varphi_{kd}^T u_i^d$

$p(r_i^x|x_i,\varphi_x) = \prod_{j=1}^{m_1}[p(r_{ij}^x = 1|x_i,\varphi_x)]^{r_{ij}^x}[1 - p(r_{ij}^x = 1|x_i,\varphi_x)]^{1-r_{ij}^x}$

$\text{logit}\{p(r_{ij}^x = 1|x_i,\varphi_x)\} = \cdots = \varphi_x^T u_i^x$

Hints for the above specified LogitReg mechanisms for binomial missingness.

(1) the conditional missing distribution can be easily be derived.
(2) not too many nuisance parameters
(3) naturally model missing proba by logistic regression.

Hints: (1) true missing mechanism is unknown. a comparison of modeling strategy can be viewed as a sensitivity analysis for model misspeification of missingness
(2) unknown parameters φ_{ky} and φ_{kd} are different across distinct components.

model comparison: diff # component / diff missing mech

$DIC = -4\mathbb{E}_{\theta*,F_m}\{\log p(F_o,F_m|\theta*)|F_o\} \approx \frac{1}{J}\sum_{j=1}^J \log p(F_o,F_m^{(j)}|\theta_*^{(j)})$

$+2\mathbb{E}_{F_m}\{\log p(F_o,F_m|\mathbb{E}_{\theta*}[\theta*|F_o,F_m])|F_o\}$

more reliable result. $\quad \tilde{\theta}_0 \leftarrow j$ loop 1 $\quad J'$ loop 2

$\mathbb{E}_{\theta*}[\theta*|F_o,F_m^{(j)}] \approx \bar{\theta}_*^{(j)} = \frac{1}{L}\sum_\ell \theta_*^{(j,\ell)}$

$\mathbb{E}_{F_m}\{\log p(F_o,F_m|\mathbb{E}_{\theta*}[\cdots|\cdots])|F_o\} \approx \frac{1}{J}\sum_j \log p(F_o,F_m^{(j)}|\bar{\theta}_*^{(j)})$

$\log p(F_o,F_m|\theta*) = \sum_{i=1}^n (\log p(y_i,\omega_i,d_i,z_i,x_i,r_i^y,r_i^d,r_i^x|\theta*)$

$\cdot\cdot = \log p(y_i|\omega_i,\mu_k,\Lambda_k,\Psi_k,z = k)$

$+ \log p(\eta_i|\xi_i,d_i,\Lambda_{\omega k},\Psi_{\delta k},z = k) + \log p(\xi_i|\Phi_k,z_i = k)$

$+ (\log p(z_i = k|\tau,x_i) + \log p(d_i|\tau_{kd},z_i = k) + \log p(x_i|\tau_x)$

$+ (\log p(r_i^y|y_i,\varphi_{ky},z_i = k) + \log p(r_i^d|d_i,\varphi_{kd},z_i = k) + \log p(r_i^x|\varphi_x,x_i)$

$= -\frac{1}{2}\{p\log(2\pi) + \log|\Psi_k| + (y_i - \mu_k - \Lambda_k\omega_i)^T\Psi_k^{-1}(\cdots)\}$

$-\frac{1}{2}\{q_1\log(2\pi) + \log|\Psi_{\delta k}| + (\eta_i - \Lambda_{\omega k}G(\omega_i))^T\Psi_{\delta k}^{-1}(\cdots)\}$

$-\frac{1}{2}\{q_2(\log(2\pi) + \log|\Phi_k| + \xi_i^T\Phi_k^{-1}\xi_i\}$

$+ [\tau_k^T x_i - \log\{\sum\exp\{\tau_k^T x_i\}\} + \log p(d_i|\cdots) + \log p(x_i|\cdots)$

$+ [(\sum_{j=1}^p r_{ij}^y)(\varphi_{ky}^T u_i^y) - p\log\{1 + \exp(\varphi_{ky}^T u_i^y)\}]$

$+ [(\sum_{j=1}^{m_2} r_{ij}^d)(\varphi_{kd}^T u_i^d) - m_2\log\{1 + \exp\{\varphi_{kd}^T u_i^d\}\}]$

$+ [(\sum_{j=1}^{m_1} r_{ij}^x)(\varphi_x^T u_i^x) - m_1\log\{1 + \exp\{\varphi_x^T u_i^x\}\}]$

Hints for Interpretation

* parameter differences among components
* necessaryness of MANR.
* necessaryness of multiple components.
* missing mechanisms differences among components.

LCM (Latent Curves Model)

initial status + change rate = trajectory over time.

$\underbrace{Y \leftarrow \eta_1,\eta_2 \quad X_k \leftarrow \xi_{1k},\xi_{2k}}_{1^{st}\text{ order latent}}$ $\quad \underbrace{u_{1:p} \leftarrow \xi_{1m},\xi_{2m}}_{2^{nd}\text{ order latent}}$.

- effects of 1^{st} and 2^{nd} latent variables on the growth factors of the outcome variables
- interaction of latent variables to assess the joint effects of dynamic latent variables.
- includes mixed continuous and ordered categorical data
- accounts for missing data.

$y = \Lambda\eta + \varepsilon_y$ $\qquad Y$ $\qquad \Lambda = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_0 & t_1 & \cdots & t \end{bmatrix}^T$

$T\times1$ $\ T\times m$ $\ m\times1$ $\ T\times1 \sim N(0,\Psi_y)$ $\quad \Psi_y = diag\{\psi_{yt}\}$ $\ t = 1,\cdots,T$.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & \cdots & 0 \\ 0 & \Lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda \end{bmatrix}\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_r \end{bmatrix} + \begin{bmatrix} \varepsilon_{x1} \\ \varepsilon_{x2} \\ \vdots \\ \varepsilon_{xr} \end{bmatrix} \quad X$$

$* x_k = \Lambda\xi_k + \varepsilon_{xk}$ $\quad k = 1,\cdots,r$

① $u_t = \Lambda_{ut}\omega_{ut} + \varepsilon_{ut}$ $\quad t = 1,\cdots,T$

$p\times1$ $\ p\times q$ $\ q\times1$ $\quad p\times1 \sim N(0,\Psi_{ut})$ $\ \Psi_{ut} = diag\{\psi_{utj}\}_{j=1}^p$

$\omega_k = (\omega_{u1k},\cdots,\omega_{uTk})^T$, $\ \varepsilon_{\omega k} = (\varepsilon_{\omega1k},\cdots,\varepsilon_{\omega Tk})^T$

② $\omega_k = \Lambda\xi_{r+k} + \varepsilon_{\omega k}$ $\quad k = 1,\cdots,q$.

* ξ_{r+k} $\ m\times1$ latent growth-factor.

$\xi = (\xi_1^T,\cdots,\xi_r^T,\xi_{r+1}^T,\cdots,\xi_{r+q}^T)^T$

$\Rightarrow \eta = \beta + \Pi\eta + \Gamma F(\xi) + \delta$.

$= \beta + \Lambda_S G(\zeta) + \delta$ $\quad \Lambda_S = (\Pi,\Gamma)$ $\quad \zeta = (\eta^T,\xi^T)^T$

LCM with ordered categorical variable.

$x_{tk} = j$ if $\alpha_{tk,j} \leq z_{tk} < \alpha_{tk,j+1}$, $\ j = 0,\cdots,h_k$. $\ k = 1,\cdots$ $\ t = 1,\cdots,T$.

identification condition.

① set $\alpha_{tk,1} = 0$ $\qquad$ for $k = 1,\cdots,r$, $t = 1,\cdots,T$.

② set $\psi_{xtk} = 1$

OrdCat $p(\alpha_{k,2},\cdots,\alpha_{k,b_k-1}) \propto C$

$p(\alpha_k,Y_k|Z_k,\theta,\Omega) = p(\alpha_k|Z_k,\theta,\Omega)p(Y_k|\alpha_k,Z_k,\theta,\Omega)$

$p(\alpha_k|Z_k,\theta,\Omega) \propto \prod_{i=1}^n[\Phi^*\{\psi_{yk}^{-\frac{1}{2}}(\alpha_{k,z_{ik}+1} - \mu_{yk} - \Lambda_{yk}^T\omega_i)\}$

$-\Phi^*\{\psi_{yk}^{-\frac{1}{2}}(\alpha_{k,z_{ik}} - \mu_{yk} - \Lambda_{yk}^T\omega_i)\}]$

$[y_{ik}|\alpha_k,Z_k,\theta,\Omega] \sim N(\mu_{yk} + \Lambda_{yk}^T\omega_i, \psi_{yk})I_{(\alpha_{k,z_{ik}},\alpha_{k,z_{ik}+1})}(y_{ik})$

EFD.

$p(\alpha_k,Y_k|Z_k,\theta,\Omega) \propto \prod_{i=1}^n\Phi^*\{\psi_{yk}^{-\frac{1}{2}}(y_{ik} - \mu_{yk} - \Lambda_{yk}^T\omega_i)\}$

$I_{(\alpha_{k,z_{ik}},\alpha_{k,z_{ik}+1})}(y_{ik})$.

EFD $\ p(\Omega|Y,\theta) \propto \prod_{i=1}^n p(y_i|\omega_i,\theta)p(\eta_i|\xi_i,\theta)p(\xi_i|\theta)$

$p(\omega_i|y_i,\theta) \propto \exp\{\sum_{k=1}^p[y_{ik}\vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\varepsilon k} -$

$\frac{1}{2}[\eta_i - Bd_i - \Pi\eta_i - \Gamma F(\xi_i)]^T\Psi_\delta^{-1}[\cdots] + \xi_i^T\Phi^{-1}\xi_i]$

$p(A_k|Y,\Omega,\Lambda_k,\psi_{\varepsilon k}) \propto \exp\{\sum_{i=1}^n\frac{y_{ik}\vartheta_{ik} - b(\vartheta_{ik})}{\psi_{\varepsilon k}} - \frac{1}{2}(A_k - A_{0k})^T H_{0k}^{-1}\cdots$

$p(\psi_{\varepsilon k}|Y,\Omega,A_k,\Lambda_k) \propto \psi_{\varepsilon k}^{-(\frac{n}{2}+\alpha_{0\varepsilon k}-1)}\exp\{\sum_i(\cdots + c_k(y_{ik},\psi_{\varepsilon k})) - \frac{\beta_{0\varepsilon k}}{\psi_{\varepsilon k}}\}$

$p(\Lambda_k|Y,\Omega,A_k,\psi_{\varepsilon k}) \propto \exp\{\sum_{i=1}^n\frac{y_{ik}\vartheta_{ik} - b(\vartheta_{ik})}{\psi_{\varepsilon k}} - \frac{1}{2}\psi_{\varepsilon k}^{-1}(\Lambda_k - \Lambda_{0k})^T H_{0yk}^{-1}\cdots)$

SEM Bayesian is same.