Bayesian Structural Equation Modeling: An Overview and Some Recent Results

Sik-Yum Lee

IMPS 2011, Hong Kong
1. INTRODUCTION

2. BAYESIAN APPROACH
   
   Estimation
   
   Model Comparison
   
   Applications

3. BAYESIAN ANALYSIS OF A LONGITUDINAL SEM

4. BAYESIAN ANALYSIS OF A NON-PARAMETRIC SEM

5. CONCLUSION
1. INTRODUCTION

Basic Structural Equation Model (SEM):

Measurement Equation:

\[ y_i = \mu + \Lambda \omega_i + \epsilon_i; \quad i = 1, \cdots, n \]

\[ \omega_i = \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \]

Structural Equation: \( \eta_i = \Pi \eta_i + \Gamma \xi_i + \delta_i \)

Regression models with latent variables.

As latent variables are random, regression methods cannot apply.
Covariance Structure Analysis: Analyze the covariance matrix of $\mathbf{y}$, $\Sigma(\theta)$, based on the sample covariance matrix $\mathbf{S}$.

Difficulties in analyzing complex SEMs and/or data structures:

For examples:

- NSEM with a nonlinear structural equation.
- Two level NSEM with ordered categorical data.
- Mixtures NSEMs, etc.
Advantages of the Bayesian Approach:

a) A more flexible approach to deal with complex situations.

b) Utilizes useful prior information (if available).

c) Achieves reliable results with small/moderate sample sizes (Scheines, Hoijtink & Boomsma, 1999; Dunson, 2000; Lee & Song, 2004).

d) Gives direct estimates of latent variables.
2. BAYESIAN APPROACH

2.1 Bayesian Estimation

\( M \): The model of interest.

\( \theta \): Vector of unknown parameters in \( M \).

\( D_0 \): Observed data; continuous and/or discrete data.

\( D_u \): Unobserved data; missing data, hidden continuous values underlying ordered categorical data, etc.

\( \Omega \): Various types of latent variables.

The main issue is to estimate \( \theta \).
Treat $\theta$ as random with prior pdf $p(\theta)$

$p(\theta|D_0)$ posterior density of $\theta$ given $D_0$, under $M:$
behavior of $\theta$ under the given data.

**Posterior Analysis**

$$p(\theta|D_0) \propto p(D_0|\theta)p(\theta)$$
$$\log p(\theta|D_0) \propto \log p(D_0|\theta) + \log p(\theta)$$
$$\propto \log \text{likelihood} + \log \text{prior}$$

Use information available from the data $D_0$, and prior information from $p(\theta)$. 
**Prior Distribution**

Non-informative prior distributions: $p(\theta)$ proportional to a constant or has an extremely large variance. No prior information available.

Informative prior distributions: Used when some prior knowledge is available.

If $p(\theta)$ is chosen, s.t. $p(\theta)$ and $p(\theta|D_0)$ are of the same form, then $p(\theta)$ is called a conjugate distribution.
Conjugate prior distributions for parameters in SEMs:

\[
\begin{align*}
\mu & \overset{D}{=} N[\mu_0, \Sigma_0] \\
\psi_{\varepsilon k} & \overset{D}{=} \text{Inverted Gamma}(\alpha_{0\varepsilon k}, \beta_{0\varepsilon k}). \\
(\text{Similarly for } \psi_{\delta}) \\
\Lambda_k & \overset{D}{=} N[\Lambda_{0k}, H_{0k}], \text{ } k^{th} \text{ row of } \Lambda \\
\text{cov}(\xi) & = \Phi \overset{D}{=} IW[R_0, \rho_0],
\end{align*}
\]

where \( \mu_0, \Sigma_0, \alpha_{0\varepsilon k}, \beta_{0\varepsilon k}, \Lambda_{0k}, H_{0k}, R_0, \rho_0 \) are hyperparameters whose values are specified based on the prior information or knowledge.
Bayesian estimate of $\theta$: The mean of $p(\theta|D_0)$.

Usually not in closed form.

Simulate a large sample of $\theta = \{\theta^{(1)}, \ldots, \theta^{(T)}\}$ from $p(\theta|D_0)$, and get the Bayesian estimate as

$$\hat{\theta} = T^{-1} \sum_{t=1}^{T} \theta^{(t)}$$

How to get this sample?
Usually, it is hard to simulate observations from $p(\theta|D_0)$

_Data Augmentation:_ Augment $D_0$ with $D_u$ and $\Omega$.

Consider the joint posterior $p(\theta, D_u, \Omega|D_0)$.

Simulate $(\theta, D_u, \Omega)$ from $p(\theta, D_u, \Omega|D_0)$ via some Markov chain Monte Carlo (MCMC) methods in statistical computing.
**Gibbs Sampler:** Begin with any starting values \( \theta^{(0)}, \ D_u^{(0)}, \ \Omega^{(0)} \).

At the \( j^{th} \) iteration with \( \theta^{(j)}, \ D_u^{(j)}, \ \Omega^{(j)} \); simulate

\[
\theta^{(j+1)} \text{ from } p(\theta|D_u^{(j)}, \Omega^{(j)}, D_0),
\]

\[
D_u^{(j+1)} \text{ from } p(D_u|\theta^{(j+1)}, \Omega^{(j)}, D_0),
\]

\[
\Omega^{(j+1)} \text{ from } p(\Omega|\theta^{(j+1)}, D_u^{(j+1)}, D_0).
\]

Need to derive various components in the full conditional distributions \( p(\theta|D_u, \Omega, D_0) \), \( p(D_u|\theta, \Omega, D_0) \), and \( p(\Omega|\theta, D_u, D_0) \).
Full conditional distributions are easier to deal with than $p(\theta|D_0)$.

i. Given $\Omega$, SEM is the regression model.

ii. Given $D_u$, e.g. the hidden continuous values, difficulties related to ordered categorical variables can be solved, etc.

Most of the full conditional distributions: Normal, Gamma, etc.

For non-standard conditional distributions, some standard MCMC methods (MH algorithm) can be used to draw observations.
Check convergence: Parallel sequences generated with starting values mixed well together.
After achieving convergence, say after $J$ iterations,
\[
(\theta^{(J+1)}, D_u^{(J+1)}, \Omega^{(J+1)}) \ldots \text{ can be regarded as observations from } p(\theta, D_u, \Omega | D_0).
\]

Collect \{|(\theta^{(t)}, D_u^{(t)}, \Omega^{(t)}), \; t = J + 1, \cdots, J + T\} \text{ for statistical inference.}

\[
\hat{\theta} = T^{-1} \sum_{t=1}^{T} \theta^{(t)}, \quad \hat{\Omega} = T^{-1} \sum_{t=1}^{T} \Omega^{(t)}.
\]

WinBUGS (Spiegelhalter et al. 2003) gives Bayesian estimates of parameters and latent variables for many SEMs.
2.2 Bayesian Model Comparison

\( D_0: \) Given data set.

\( M_1, M_2: \) Competing SEMs with \( \theta_1 \) and \( \theta_2 \), respectively.

(i) Bayes Factor: \( B_{10} = \frac{p(D_0|M_1)}{p(D_0|M_0)} \).

Select \( M_0 \) if \( B_{10} < 1 \);

Select \( M_1 \) if \( B_{10} > 1 \).


\[
p(D_0|M_k) = \int p(\theta_k, D_0|M_k) \, d\theta_k = \int p(D_0|\theta_k, M_k) \, p(\theta_k) \, d\theta_k.\]

Computed via path sampling (see, Lee, 2007).
(ii) Deviance Information Criterion (DIC, Spiegelhalter et al., 2002):

\[
DIC_k = -2 \frac{1}{T} \sum_{t=1}^{T} \log p(D_0|\theta_k^{(t)}, M_k) + 2d_k
\]

\[d_k = \text{dimension of } \theta_k\]

Model with smaller DIC value is selected.

For many SEMs, DIC values are available in WinBUGS.
(iii) $L_\nu$ Measure:

Let $D_0^{rep} = \{D_1^{rep}, \ldots, D_n^{rep}\}$ be a response from a replicate experiment having from the same $p(D_0|\theta)$. The $L_\nu$ measure (Ibrahim, Chen & Sinha, 2001) is defined by

$$
L_\nu(D_0) = \sum_{i=1}^{n} \text{tr}\left\{ \text{cov}(D_i^{rep}|D_0) \right\} \\
+ \nu \sum_{i=1}^{n} \text{tr}\left[ \left\{ E(D_i^{rep}|D_0) - D_i \right\} \left\{ E(D_i^{rep}|D_0) - D_i \right\}' \right]
$$

Expectation is taken w.r.t. $D_0^{rep}|D_0$, computed from simulated observations in estimation.

Model with smaller $L_\nu$ is selected.

$\nu$ usually taken as 0.5.

See Song et al. (2011) for more details and some applications.
2.4 Applications to SEMs or Related Models

a. SEMs with mixed continuous and discrete data:
   Shi & Lee (2000, JRSSB), Dunson (2000, JRSSB),
   van Onna (2002, Psym), Dunson & Herring (2005, Biostatistics),
   Song et al. (2009, Stat. Med.)

b. SEMs with missing data (MAR/Non-ignorable):

c. Nonlinear SEMs:
   Arminger & Muthen (1997, Psym), Lee & Song (2003, Psym)

d. Multilevel SEMs:
   Ansari & Jedidi (2000, Psym),
e. Mixture SEMs:

Jedidi, Jagpal & DeSarbo (1997, Marketing Sci.),

f. Item Response Models:

Begnin & Glas (2001, Psym), Fox & Glas (2001, Psym),
Miyazaki & Hoshino (2009, Psym)

g. Bayesian Semiparametric SEMs:

h. Longitudinal SEMs:
   Dunson (2003, JASA), Song, et al. (2011, SEM)

i. SEM with a Non-parametric Structural Equation:
   Song & Lu (2010, J. Comp. Graphical Stat.)

j. Transformation SEMs:
   Song & Lu (2011)

*Data Augmentation, MCMC methods*
3. BAYESIAN ANALYSIS OF A LONGITUDINAL SEM

3.1 Motivation

\[
\begin{array}{cccc}
SEM_1 & SEM_2 & \cdots & SEM_T \\
\hline
t = 1 & t = 2 & \cdots & t = T \\
\end{array}
\]

For investigating various behaviors that are

i. invariant over time,

ii. dynamically change over time.
3.2 The Model

Let \( u_{gt} \) be an observed random vector for the \( g^{th} \) \((g = 1, \cdots, G)\) individual measured at time point \( t \) \((t = 1, \cdots, T)\). A two-level model for \( u_{gt} \) is defined by: (Song et al., 2010)

\[
  u_{gt} = y_g + v_{gt},
\]  

\( y_g \): second level random vector (independent of \( t \)) for accounting characteristics invariant with \( t \)

\( v_{gt} \): first level random vector for accounting characteristics that are changed dynamically with \( t \)
The Second level model

For assessing invariant characteristics over time:

\[ y_g = A_0 c_{g0} + \Lambda_0 \omega_{g0} + \varepsilon_{g0}, \quad g = 1, \ldots, G \]  

(2)

c_{g0}: vector of covariates

\omega_{g0}: vector of latent variables

\varepsilon_{g0}: vector of residual errors

A_0 and \Lambda_0: matrices of unknown coefficients

\omega_{g0} is i.i.d \ N[0, \Phi_0]

\varepsilon_{g0} is independent of \omega_{g0}, and i.i.d \ N[0, \Psi_0], \quad \Psi_0 \text{ is diagonal}

Not depending on \( t \), invariant over time
The first-level dynamic model

For $g = 1, \cdots, G, \ t = 1, \cdots, T$, the measurement equation is:

$$v_{gt} = A_{t}c_{gt} + \Lambda_{t}\omega_{gt} + \varepsilon_{gt},$$  \hspace{1cm} (3)

where the definitions of $A_{t}, \ c_{gt}, \ \Lambda_{t}, \ \omega_{gt},$ and $\varepsilon_{gt}$ are similar to those given in equation (2), except here they are defined at time $t$ nested within the individual $g$.

Here, $\varepsilon_{gt}$ are independently dist. as $N[0, \Psi_t]$, and $\Psi_t$ is diagonal.
Let $\omega_{gt} = (\eta_{gt}', \xi_{gt}')'$. The relationships among $\eta_{gt}$ and $\xi_{gt}$, covariates, and latent vectors at previous times are studied through the following structural equation:

$$
\eta_{gt} = B_0 d_{g0} + B_t d_{gt} + \Gamma_t F_t(\eta_{g1}, \cdots, \eta_{g,t-1}, \xi_{g1}, \cdots, \xi_{g,t-1}, \xi_{gt}) + \delta_{gt},
$$

(4)

$B_0$, $B_t$, $\Gamma_t$: matrices of unknown coefficients

d$_{g0}$: covariates that are invariant with time

d$_{gt}$: covariates variant with time

$F_t$: differentiable vector-valued function of $\xi_{gt}$ at the current time $t$, and $\eta_{g1}, \cdots, \eta_{g,t-1}, \xi_{g1}, \cdots, \xi_{g,t-1}$ at previous times

$\delta_{gt}$: residual errors, independently distributed as $N[0, \Psi_{\delta_t}]$

$\xi_g = (\xi_{g1}', \cdots, \xi_{g,T}') \overset{D}{=} N[0, \Phi]$
Some simple examples of $F_t(\eta_{g1}, \cdots, \eta_{g,t-1}, \xi_{g1}, \cdots, \xi_{g,t-1}, \xi_{gt})$:

For $t = 1, \cdots, T$, $\omega_{gt} = (\eta_{gt}, \xi_{1gt}, \xi_{2gt})'$,

$$\eta_{gt} = \gamma_1 \eta_{g1} + \cdots + \gamma_{t-1} \eta_{g,t-1} + \gamma_1 \xi_{1gt} + \gamma_2 \xi_{2gt} + \gamma_3 \xi_{1gt} \xi_{2gt} + \delta_{gt},$$

$$\eta_{gt} = \gamma_1 \eta_{g,t-1} + \gamma_2 \xi_{1g,t-1} + \gamma_3 \xi_{2g,t-1} + \gamma_4 \xi_{1gt} + \gamma_5 \xi_{2gt} + \gamma_6 \xi_{1gt}^2$$

$$+ \gamma_7 \xi_{2gt}^2 + \gamma_8 \xi_{1gt} \xi_{2gt} + \delta_{gt}.$$

May add covariates.
Let $\mu_{gt} = \{X_{gt}, Z_{gt}\}$, $X_{gt}$ continuous, $Z_{gt}$ ordered categorical.

$D_0 = (X_0, Z_0)$, observed continuous data $X_0$,

observed ordered categorical data $Z_0$.

$D_u = \{\text{missing data, hidden continuous values underlying } Z_0\}$.

$\Omega_g = \{y_1, \cdots, y_G\}$,

$\Omega_0 = \{\omega_{10}, \cdots, \omega_{G0}\}$,

$\Omega_1 = \{\omega_1, \cdots, \omega_G\}$, $\omega_g = \{\omega_{g1}, \cdots, \omega_{gT}\}$, $\omega_{gt} = \begin{bmatrix} \eta_{gt} \\ \xi_{gt} \end{bmatrix}$,

$\theta = \left[ \{A_0, B_0, \Lambda_0, \Phi_0, \Psi_0\}, \right.$

$\left. \{(A_t, \Lambda_t, \Psi_t, B_t, \Gamma_t), \ t = 1, \cdots, T\}, \Phi \right]$,

$\alpha = \text{unknown thresholds}$.
Data Augmentation: Augment $D_0$ with $D_u$, $\Omega_g$, $\Omega_0$, $\Omega_1$.

Consider $p(\theta, \alpha, D_u, \Omega_g, \Omega_0, \Omega_1 | D_0)$.

Gibbs sampler: Simulate:

\[
\begin{align*}
p(\alpha, D_u | \theta, \Omega_g, \Omega_0, \Omega_1, D_0) \\
p(\theta | \alpha, D_u, \Omega_g, \Omega_0, \Omega_1, D_0) \\
p(\Omega_g | \theta, \alpha, D_u, \Omega_0, \Omega_1, D_0) \\
p(\Omega_0 | \theta, \alpha, D_u, \Omega_g, \Omega_1, D_0) \\
p(\Omega_1 | \theta, \alpha, D_u, \Omega_g, \Omega_0, D_0)
\end{align*}
\]

See Song et al. (2010) for technical details and numerical results.
4. BAYESIAN ANALYSIS OF A SEM WITH A NON-PARAMETRIC STRUCTURAL EQUATION

4.1 Motivation

For SEMs, model with a given specific parametric structural equation may not be realistic.

Consider: $\eta_1 = ax_i + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \delta_i$

Regression — observed variables — data
SEM — latent variables — ?

Desirable to develop model with its structural equation formulated through unspecific function of latent variables and covariates.
4.2 The Model (Song & Lu, 2010)

Measurement Equation: \( y_i = A c_i + \Lambda \omega_i + \varepsilon_i, \; i = 1, \cdots, n \) (5)

\[ \omega_i = \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix}, \quad \eta_i : \; r \times 1 \text{ vector of outcome l.v.} \]
\[ \xi_i : \; s \times 1 \text{ vector of explanatory l.v.} \sim \mathcal{N}[0, \Phi] \]

Structural Equation: for the \( j^{th} \) element \( \eta_{ij} \) in \( \eta_i \): For \( j = 1, \cdots, r \):

\[ \eta_{ij} = g_{j1}(x_{i1}) + \cdots + g_{jd}(x_{id}) + f_{j1}(\xi_{i1}) + \cdots + f_{js}(\xi_{is}) + \delta_j, \] (6)

\( x_{i1}, \cdots, x_{id} : \) covariates

\( g_{j1}, \cdots, g_{jd} : \) unspecified functions

\( f_{j1}, \cdots, f_{js} : \) unspecified functions
For any unspecified general function $f(x)$ with a continuous $2^{nd}$ order derivative, it can be modeled by a sum of B-splines basis with $K$ knots in the domain of $x$ as (see De Boor, 1978):

$$f(x) = \sum_{k=1}^{K} \beta_k B_k(x),$$

$K$: number of knots (usually 10-30)

$\beta_k$: unknown parameter

$B_k(x)$: B-spline of appropriate order

See examples B-splines in the text book (De Boor, 1978)
The basic idea for modeling (6) is: (suppress subscript \( j \))

\[
\eta_i = \sum_{k=1}^{K_{b_1}} b_{1k} B_{1_{1k}}^x (x_{i1}) + \cdots + \sum_{k=1}^{K_{b_d}} b_{dk} B_{dk}^x (x_{id}) \\
+ \sum_{k=1}^{K_1} \beta_{1k} B_{1k} (\xi_{i1}) + \cdots + \sum_{k=1}^{K_s} \beta_{sk} B_{sk} (\xi_{is}) + \delta.
\]

(7)

\( K \)'s: number of knots

\( b_{1k}, \cdots, b_{dk_{1b}}, \cdots, \cdots, b_{dk_{bd}} \): unknown parameters

\( \beta_{1k}, \cdots, \beta_{1K_1}, \cdots, \cdots, \beta_{sk_{sK}} \): unknown parameters

\( B_{mk}^x, B_{mk} \): modified B-splines
Let
\[ \mathbf{Y} = (y_1, \ldots, y_n) \]
\[ \Omega = (\omega_1, \ldots, \omega_n) \]

\( \theta \): vector consists of all unknown parameters

**Posterior Analysis:** Consider the joint posterior distribution with conjugate priors. Apply Gibbs sampler (with MH algorithm):

a) draw \( \Omega \) from \( p(\Omega | \theta, \mathbf{Y}) \)

b) draw \( \theta \) from \( p(\theta | \Omega, \mathbf{Y}) \)

Only rough ideas.

See Song & Lu (2010) for much more details in formulating the model and solving the difficulties in the posterior analysis.
4.3 A Simulation Study

Measurement equation:

\[ y_i = a + \Lambda \omega_i + \varepsilon_i, \]

where \( a = (a_1, \ldots, a_{12})' \), \( \omega_i = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3})' \), and

\[
\Lambda' = \begin{bmatrix}
1 & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{83} & \lambda_{93} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{11,4} & \lambda_{12,4} \\
\end{bmatrix},
\]

where \( \xi \overset{\mathcal{D}}{=} N[0, \Phi] \), and

\[ a_j = 0.5, \ \psi_j = 0.3, \ \text{for } j = 1, \ldots, 12 \]

\[ \lambda_{21} = \lambda_{31} = \lambda_{52} = \lambda_{62} = \lambda_{83} = \lambda_{93} = \lambda_{11,4} = \lambda_{12,4} = 0.8, \]

\[ \phi_{ii} = 1.0, \ \phi_{ik} = 0.2, \ \text{for all } i, k, \ i \neq k; \ \psi_\delta = 0.3 \]
Structural equation:

\[ \eta_i = g(x_i) + f_1(\xi_{i1}) + f_2(\xi_{i2}) + f_3(\xi_{i3}) + \delta_i, \]

where

\[ g(x) = (x/2)^3 \]

\[ f_1(\xi) = \sin(1.5\xi) - \xi \]

\[ f_2(\xi) = 1.65 - \exp(\xi) \]

\[ f_3(\xi) = -0.5 + \exp(2\xi)/[1 + \exp(2\xi)] \]
$n = 300, \ K = 20, \ number\ of\ replications=100.$

<table>
<thead>
<tr>
<th>Par.</th>
<th>True</th>
<th>Bias</th>
<th>RMS</th>
<th>Par.</th>
<th>True</th>
<th>Bias</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{21}$</td>
<td>0.8</td>
<td>-0.001</td>
<td>0.018</td>
<td>$\phi_{11}$</td>
<td>1.0</td>
<td>-0.057</td>
<td>0.110</td>
</tr>
<tr>
<td>$\lambda_{31}$</td>
<td>0.8</td>
<td>-0.001</td>
<td>0.019</td>
<td>$\phi_{12}$</td>
<td>0.2</td>
<td>-0.022</td>
<td>0.070</td>
</tr>
<tr>
<td>$\lambda_{52}$</td>
<td>0.8</td>
<td>0.022</td>
<td>0.044</td>
<td>$\phi_{13}$</td>
<td>0.2</td>
<td>-0.010</td>
<td>0.064</td>
</tr>
<tr>
<td>$\lambda_{62}$</td>
<td>0.8</td>
<td>0.021</td>
<td>0.048</td>
<td>$\phi_{22}$</td>
<td>1.0</td>
<td>-0.037</td>
<td>0.113</td>
</tr>
<tr>
<td>$\lambda_{83}$</td>
<td>0.8</td>
<td>0.022</td>
<td>0.050</td>
<td>$\phi_{23}$</td>
<td>0.2</td>
<td>-0.003</td>
<td>0.058</td>
</tr>
<tr>
<td>$\lambda_{93}$</td>
<td>0.8</td>
<td>0.023</td>
<td>0.048</td>
<td>$\phi_{33}$</td>
<td>1.0</td>
<td>-0.025</td>
<td>0.116</td>
</tr>
<tr>
<td>$\lambda_{11,4}$</td>
<td>0.8</td>
<td>0.007</td>
<td>0.049</td>
<td>$\psi_{\delta}$</td>
<td>0.3</td>
<td>0.020</td>
<td>0.040</td>
</tr>
<tr>
<td>$\lambda_{12,4}$</td>
<td>0.8</td>
<td>0.016</td>
<td>0.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The solid curves represent the true curves, the dashed and dot-dashed curves respectively represent the estimated posterior means and the 5%- and 95%-pointwise quantiles on the basis of 100 replications.
Remarks:

- The estimated curves correctly capture the true functional relationships.
- The parameter estimates are accurate: all bias close to zero, most RMS values are below 0.05.
- The sensitivity analysis shows that the Bayesian results are robust to different prior inputs.
- For analysis of some real-world data sets, see Song & Lu (2010).
5. CONCLUSION

a) The Bayesian approach with data augmentation and MCMC methods is flexible for analyzing complex SEMs.

b) Future research/work

(i) Dynamic SEM with non-ignorable missing data

(ii) SEMs with structural equation involving nonparametric function of interaction of latent variables; such as

\[ \eta_i = f_1(\xi_{i1}) + f_2(\xi_{i2}) + f_3(\xi_{i1}\xi_{i2}). \]

(iii) Dynamic SEM with nonparametric structural equation.

(iv) Development of use-friendly software for applied researchers.
Thank you!