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:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i; \quad i = 1, \cdots, n$$

$$\boldsymbol{\omega}_i = \begin{pmatrix} \boldsymbol{\eta}_i \\ \boldsymbol{\xi}_i \end{pmatrix}$$

Structural Equation: $\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i$

Regression models with latent variables.

As latent variables are random, regression methods cannot apply.

Covariance Structure Analysis: Analyze the covariance matrix of $\mathbf{y}$, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, 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.

$\boldsymbol{\theta}$: Vector of unknown parameters in M.

$\mathbf{D}_0$: Observed data; continuous and/or discrete data.

$\mathbf{D}_u$: Unobserved data; missing data, hidden continuous values underlying ordered categorical data, etc.

$\boldsymbol{\Omega}$: Various types of latent variables.

The main issue is to estimate $\boldsymbol{\theta}$.

Treat $\boldsymbol{\theta}$ as random with prior pdf $p(\boldsymbol{\theta})$

$p(\boldsymbol{\theta}|\mathbf{D}_0)$ posterior density of $\boldsymbol{\theta}$ given $\mathbf{D}_0$, under M:
behavior of $\boldsymbol{\theta}$ under the given data.

Posterior Analysis

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{D}_0) &\propto p(\mathbf{D}_0|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ \log p(\boldsymbol{\theta}|\mathbf{D}_0) &\propto \log p(\mathbf{D}_0|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) \\ &\propto \log \textit{likelihood} + \log \textit{prior} \end{aligned}$$

Use information available from the data $\mathbf{D}_0$, and prior information from $p(\boldsymbol{\theta})$.

Prior Distribution

Non-informative prior distributions: $p(\boldsymbol{\theta})$ 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 $p(\boldsymbol{\theta})$ is chosen, s.t. $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\mathbf{D}_0)$ are of the same form, then $p(\boldsymbol{\theta})$ is called a conjugate distribution.

Conjugate prior distributions for parameters in SEMs:

$$\begin{aligned}
\boldsymbol{\mu} &\stackrel{D}{=} N[\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0] \\
\Psi_{\varepsilon k} &\stackrel{D}{=} \text{Inverted Gamma}(\alpha_{0\varepsilon k}, \beta_{0\varepsilon k}). \\
& \quad \text{(Similarly for } \Psi_\delta\text{)} \\
\boldsymbol{\Lambda}_k &\stackrel{D}{=} N[\boldsymbol{\Lambda}_{0k}, \mathbf{H}_{0k}], \quad k^{th} \text{ row of } \boldsymbol{\Lambda} \\
\text{cov}(\boldsymbol{\xi}) = \ \boldsymbol{\Phi} &\stackrel{D}{=} IW[\mathbf{R}_0, \rho_0],
\end{aligned}$$

where $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, $\alpha_{0\varepsilon k}$, $\beta_{0\varepsilon k}$, $\boldsymbol{\Lambda}_{0k}$, $\mathbf{H}_{0k}$, $\mathbf{R}_0$, ρ_0 are hyperparameters whose values are specified based on the prior information or knowledge.

Bayesian estimate of $\boldsymbol{\theta}$: The mean of $p(\boldsymbol{\theta}|\mathbf{D}_0)$.

Usually not in closed form.

Simulate a large sample of $\boldsymbol{\theta} = \{\boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(T)}\}$ from $p(\boldsymbol{\theta}|\mathbf{D}_0)$, and get the Bayesian estimate as

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

How to get this sample?

Usually, it is hard to simulate observations from $p(\boldsymbol{\theta}|\mathbf{D}_0)$

Data Augmentation: Augment $\mathbf{D}_0$ with $\mathbf{D}_u$ and $\boldsymbol{\Omega}$.

Consider the joint posterior $p(\boldsymbol{\theta}, \mathbf{D}_u, \boldsymbol{\Omega}|\mathbf{D}_0)$.

Simulate $(\boldsymbol{\theta}, \mathbf{D}_u, \boldsymbol{\Omega})$ from $p(\boldsymbol{\theta}, \mathbf{D}_u, \boldsymbol{\Omega}|\mathbf{D}_0)$ via some Markov chain Monte Carlo (MCMC) methods in statistical computing.

Gibbs Sampler: Begin with any starting values $\boldsymbol{\theta}^{(0)},\ \mathbf{D}_u^{(0)},\ \boldsymbol{\Omega}^{(0)}$.

At the j^{th} iteration with $\boldsymbol{\theta}^{(j)},\ \mathbf{D}_u^{(j)},\ \boldsymbol{\Omega}^{(j)}$; simulate

$$\boldsymbol{\theta}^{(j+1)} \text{ from } p(\boldsymbol{\theta}|\mathbf{D}_u^{(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{D}_0),$$

$$\mathbf{D}_u^{(j+1)} \text{ from } p(\mathbf{D}_u|\boldsymbol{\theta}^{(j+1)}, \boldsymbol{\Omega}^{(j)}, \mathbf{D}_0),$$

$$\boldsymbol{\Omega}^{(j+1)} \text{ from } p(\boldsymbol{\Omega}|\boldsymbol{\theta}^{(j+1)}, \mathbf{D}_u^{(j+1)}, \mathbf{D}_0).$$

Need to derive various components in the full conditional distributions $p(\boldsymbol{\theta}|\mathbf{D}_u, \boldsymbol{\Omega}, \mathbf{D}_0)$, $p(\mathbf{D}_u|\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{D}_0)$, and $p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \mathbf{D}_u, \mathbf{D}_0)$.

Full conditional distributions are easier to deal with than $p(\boldsymbol{\theta}|\mathbf{D}_0)$.

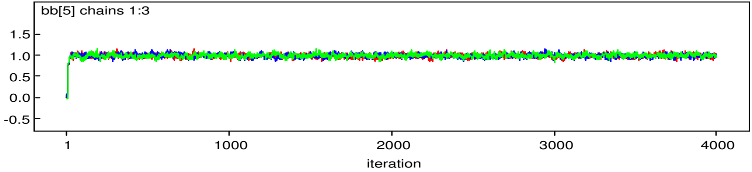
i. Given $\boldsymbol{\Omega}$, SEM is the regression model.

ii. Given $\mathbf{D}_u$, e.g. the hidden continuous values,
difficulties related to ordered categorical variables can be solved, etc.

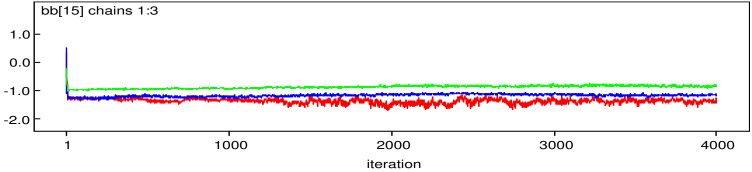
Most of the full conditional distributions: Normal, Gamma, etc.

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.



a



b

After achieving convergence, say after J iterations, $(\boldsymbol{\theta}^{(J+1)}, \mathbf{D}_u^{(J+1)}, \boldsymbol{\Omega}^{(J+1)}) \cdots$ can be regarded as observations from $p(\boldsymbol{\theta}, \mathbf{D}_u, \boldsymbol{\Omega} | \mathbf{D}_0)$.

Collect $\{(\boldsymbol{\theta}^{(t)}, \mathbf{D}_u^{(t)}, \boldsymbol{\Omega}^{(t)}),\ t = J+1, \cdots, J+T\}$ for statistical inference.

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

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

2.2 Bayesian Model Comparison

$\mathbf{D}_0$: Given data set.

$M_1, \ M_2$: Competing SEMs with $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively.

(i) Bayes Factor: $B_{10} = \frac{p(\mathbf{D}_0|M_1)}{p(\mathbf{D}_0|M_0)}$.

Select M_0 if $B_{10} < 1$;

Select M_1 if $B_{10} > 1$.

See Kass & Raftery (1995) for more detailed interpretation.

$$p(\mathbf{D}_0|M_k) = \int p(\boldsymbol{\theta}_k, \mathbf{D}_0|M_k)d\boldsymbol{\theta}_k = \int p(\mathbf{D}_0|\boldsymbol{\theta}_k, M_k)p(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k.$$

Computed via path sampling (see, Lee, 2007).

(ii) Deviance Information Criterion (DIC, Spiegelhalter et al., 2002):

$$DIC_k = -\frac{2}{T}\sum_{t=1}^{T} \log p(\mathbf{D}_0|\boldsymbol{\theta}_k^{(t)}, M_k) + 2d_k$$

$$d_k = \text{dimension of } \boldsymbol{\theta}_k$$

Model with smaller DIC value is selected.

For many SEMs, DIC values are available in WinBUGS.

(iii) L_ν Measure:

Let $\mathbf{D}_0^{\text{rep}} = \{\mathbf{D}_1^{\text{rep}}, \cdots, \mathbf{D}_n^{\text{rep}}\}$ be a response from a replicate experiment having from the same $p(\mathbf{D}_0|\boldsymbol{\theta})$. The L_ν measure (Ibrahim, Chen & Sinha, 2001) is defined by

$$L_\nu(\mathbf{D}_0) = \sum_{i=1}^{n} \text{tr}\Big\{\text{cov}(\mathbf{D}_i^{\text{rep}}|\mathbf{D}_0)\Big\} + \nu \sum_{i=1}^{n} \text{tr}\Big[\Big\{E(\mathbf{D}_i^{\text{rep}}|\mathbf{D}_0) - \mathbf{D}_i\Big\}\Big\{E(\mathbf{D}_i^{\text{rep}}|\mathbf{D}_0) - \mathbf{D}_i\Big\}'\Big]$$

Expectation is taken w.r.t. $\mathbf{D}_0^{\text{rep}}|\mathbf{D}_0$, computed from simulated observations in estimation.

Model with smaller L_ν is selected.

ν 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):
   Song & Lee (2002, Psym), Song & Lee (2009, Stat. Med.)

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

d. Multilevel SEMs:
   Ansari & Jedidi (2000, Psym),
   Ansari, Jedidi & Duke (2002, Psym), Song & Lee (2004, BJMSP)

e. Mixture SEMs:

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

Zhu & Lee (2001, Psym), Cai, Song & Hser (2010, Stat. Med.)

f. Item Response Models:

Hoijtink & Molenaar (1997, Psym), Patz & Junker (1997, JEBS),

Begnin & Glas (2001, Psym), Fox & Glas (2001, Psym),

Miyazaki & Hoshino (2009, Psym)

g. Bayesian Semiparametric SEMs:

Lee, Lu & Song (2007, Stat. Med.),

Song, Lee & Xia (2009, Stat. Med.), Yang & Dunson (2010, Psym)

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

$$\begin{array}{cccc} \text{SEM}_1 & \text{SEM}_2 & \cdots & \text{SEM}_T \\ t=1 & t=2 & \cdots & t=T \end{array}$$

For investigating various behaviors that are

i. invariant over time,

ii. dynamically change over time.

3.2 The Model

Let $\mathbf{u}_{gt}$ 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 $\mathbf{u}_{gt}$ is defined by: (Song et al., 2010)

$$\mathbf{u}_{gt} = \mathbf{y}_g + \mathbf{v}_{gt}, \tag{1}$$

$\mathbf{y}_g$: second level random vector (independent of t) for accounting characteristics invariant with t

$\mathbf{v}_{gt}$: first level random vector for accounting characteristics that are changed dynamically with t

The Second level model

For assessing invariant characteristics over time:

$$\mathbf{y}_g = \mathbf{A}_0 \mathbf{c}_{g0} + \boldsymbol{\Lambda}_0 \boldsymbol{\omega}_{g0} + \boldsymbol{\varepsilon}_{g0}, \quad g = 1, \cdots, G \tag{2}$$

$\mathbf{c}_{g0}$: vector of covariates

$\boldsymbol{\omega}_{g0}$: vector of latent variables

$\boldsymbol{\varepsilon}_{g0}$: vector of residual errors

$\mathbf{A}_0$ and $\boldsymbol{\Lambda}_0$: matrices of unknown coefficients

$\boldsymbol{\omega}_{g0}$ is i.i.d $N[\mathbf{0}, \boldsymbol{\Phi}_0]$

$\boldsymbol{\varepsilon}_{g0}$ is independent of $\boldsymbol{\omega}_{g0}$, and i.i.d $N[\mathbf{0}, \boldsymbol{\Psi}_0]$, $\quad \boldsymbol{\Psi}_0$ 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:

$$\mathbf{v}_{gt} = \mathbf{A}_t \mathbf{c}_{gt} + \boldsymbol{\Lambda}_t \boldsymbol{\omega}_{gt} + \boldsymbol{\varepsilon}_{gt}, \tag{3}$$

where the definitions of $\mathbf{A}_t,\ \mathbf{c}_{gt},\ \boldsymbol{\Lambda}_t,\ \boldsymbol{\omega}_{gt}$, and $\boldsymbol{\varepsilon}_{gt}$ are similar to those given in equation (2), except here they are defined at time t nested within the individual g.

Here, $\boldsymbol{\varepsilon}_{gt}$ are independently dist. as $N[0, \boldsymbol{\Psi}_t]$, and $\boldsymbol{\Psi}_t$ is diagonal.

Let $\boldsymbol{\omega}_{gt} = (\boldsymbol{\eta}'_{gt}, \boldsymbol{\xi}'_{gt})'$. The relationships among $\boldsymbol{\eta}_{gt}$ and $\boldsymbol{\xi}_{gt}$, covariates, and latent vectors at previous times are studied through the following structural equation:

$$\boldsymbol{\eta}_{gt} = \mathbf{B}_0\mathbf{d}_{g0} + \mathbf{B}_t\mathbf{d}_{gt} + \boldsymbol{\Gamma}_t\mathbf{F}_t(\boldsymbol{\eta}_{g1}, \cdots, \boldsymbol{\eta}_{g,t-1}, \boldsymbol{\xi}_{g1}, \cdots, \boldsymbol{\xi}_{g,t-1}, \boldsymbol{\xi}_{gt}) + \boldsymbol{\delta}_{gt}, \tag{4}$$

$\mathbf{B}_0,\ \mathbf{B}_t,\ \boldsymbol{\Gamma}_t$: matrices of unknown coefficients

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

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

$\mathbf{F}_t$: differentiable vector-valued function of $\boldsymbol{\xi}_{gt}$ at the current time t, and $\boldsymbol{\eta}_{g1}, \cdots, \boldsymbol{\eta}_{g,t-1}, \boldsymbol{\xi}_{g1}, \cdots, \boldsymbol{\xi}_{g,t-1}$ at previous times

$\boldsymbol{\delta}_{gt}$: residual errors, independently distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_{\delta t}]$

$\boldsymbol{\xi}_g = (\boldsymbol{\xi}'_{g1}, \cdots, \boldsymbol{\xi}'_{gT})' \stackrel{D}{=} N[\mathbf{0}, \boldsymbol{\Phi}]$

Some simple examples of $\mathbf{F}_t(\boldsymbol{\eta}_{g1}, \cdots, \boldsymbol{\eta}_{g,t-1}, \boldsymbol{\xi}_{g1}, \cdots, \boldsymbol{\xi}_{g,t-1}, \boldsymbol{\xi}_{gt})$:

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

$$\eta_{gt} = \gamma_{\eta 1}\eta_{g1} + \cdots + \gamma_{\eta,t-1}\eta_{g,t-1} + \gamma_{t1}\xi_{1gt} + \gamma_{t2}\xi_{2gt} + \gamma_{t3}\xi_{1gt}\xi_{2gt} + \delta_{gt},$$

$$\begin{aligned}\eta_{gt} =& \gamma_1\eta_{g,t-1} + \gamma_2\xi_{1g,t-1} + \gamma_3\xi_{2g,t-1} + \gamma_4\xi_{1gt} + \gamma_5\xi_{2gt} + \gamma_6\xi_{1gt}^2 \\ &+ \gamma_7\xi_{2gt}^2 + \gamma_8\xi_{1gt}\xi_{2gt} + \delta_{gt}.\end{aligned}$$

May add covariates.

Let $\boldsymbol{\mu}_{gt} = \{\mathbf{X}_{gt}, \mathbf{Z}_{gt}\}$, $\mathbf{X}_{gt}$ continuous, $\mathbf{Z}_{gt}$ ordered categorical.

$\mathbf{D}_0 = (\mathbf{X}_0, \mathbf{Z}_0)$, observed continuous data $\mathbf{X}_0$,

observed ordered categorical data $\mathbf{Z}_0$.

$\mathbf{D}_u$=\{missing data, hidden continuous values underlying $\mathbf{Z}_0$\}.

$$\boldsymbol{\Omega}_g = \{\mathbf{y}_1, \cdots, \mathbf{y}_G\},$$

$$\boldsymbol{\Omega}_0 = \{\boldsymbol{\omega}_{10}, \cdots, \boldsymbol{\omega}_{G0}\},$$

$$\boldsymbol{\Omega}_1 = \{\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_G\}, \ \boldsymbol{\omega}_g = \{\omega_{g1}, \cdots, \omega_{gT}\}, \ \boldsymbol{\omega}_{gt} = \begin{bmatrix} \boldsymbol{\eta}_{gt} \\ \boldsymbol{\xi}_{gt} \end{bmatrix},$$

$$\boldsymbol{\theta} = \Big[\big\{\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{\Lambda}_0, \boldsymbol{\Phi}_0, \boldsymbol{\Psi}_0\big\}, \\ \big\{(\mathbf{A}_t, \boldsymbol{\Lambda}_t, \boldsymbol{\Psi}_t, \mathbf{B}_t, \boldsymbol{\Gamma}_t), \ t = 1, \cdots, T\big\}, \ \boldsymbol{\Phi}\Big],$$

$\boldsymbol{\alpha} =$ unknown thresholds.

Data Augmentation: Augment $\mathbf{D}_0$ with $\mathbf{D}_u,\ \boldsymbol{\Omega}_g,\ \boldsymbol{\Omega}_0,\ \boldsymbol{\Omega}_1$.

Consider $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{D}_u, \boldsymbol{\Omega}_g, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1 | \mathbf{D}_0)$.

Gibbs sampler: Simulate:

$$p(\boldsymbol{\alpha}, \mathbf{D}_u | \boldsymbol{\theta}, \boldsymbol{\Omega}_g, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \mathbf{D}_0)$$

$$p(\boldsymbol{\theta} | \boldsymbol{\alpha}, \mathbf{D}_u, \boldsymbol{\Omega}_g, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \mathbf{D}_0)$$

$$p(\boldsymbol{\Omega}_g | \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{D}_u, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \mathbf{D}_0)$$

$$p(\boldsymbol{\Omega}_0 | \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{D}_u, \boldsymbol{\Omega}_g, \boldsymbol{\Omega}_1, \mathbf{D}_0)$$

$$p(\boldsymbol{\Omega}_1 | \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{D}_u, \boldsymbol{\Omega}_g, \boldsymbol{\Omega}_0, \mathbf{D}_0)$$

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)

$$\textit{Measurement Equation}: \ \mathbf{y}_i = \mathbf{A}\mathbf{c}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, \ i = 1, \cdots, n \quad (5)$$

$$\boldsymbol{\omega}_i = \begin{pmatrix} \boldsymbol{\eta}_i \\ \boldsymbol{\xi}_i \end{pmatrix}, \qquad \begin{array}{ll} \boldsymbol{\eta}_i: & r \times 1 \text{ vector of outcome l.v.} \\ \boldsymbol{\xi}_i: & s \times 1 \text{ vector of explanatory l.v.} \overset{D}{=} N[\mathbf{0}, \boldsymbol{\Phi}] \end{array}$$

Structural Equation: for the j^{th} element η_{ij} in $\boldsymbol{\eta}_i$: For $j = 1, \cdots, r$:

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

$x_{i1}, \cdots, x_{id}$: 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)

β_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)

$$\begin{aligned}\eta_i = &\sum_{k=1}^{K_{b_1}} b_{1k} B^x_{1k}(x_{i1}) + \cdots + \sum_{k=1}^{K_{b_d}} b_{dk} B^x_{dk}(x_{id}) \\ &+ \sum_{k=1}^{K_1} \beta_{1k} B_{1k}(\xi_{i1}) + \cdots + \sum_{k=1}^{K_s} \beta_{sk} B_{sk}(\xi_{is}) + \delta.\end{aligned} \tag{7}$$

K's: number of knots

$b_{1k}, \cdots, b_{dK_{1b}}, \cdots, \cdots, b_{dK_{bd}}$: unknown parameters

$\beta_{1k}, \cdots, \beta_{1K_1}, \cdots, \cdots, \beta_{sK_s}$: unknown parameters

B^x_{mk}, B_{mk}: modified B-splines

Let

$$\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$$

$$\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$$

$\boldsymbol{\theta}$: vector consists of all unknown parameters

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

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

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

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:

$$\mathbf{y}_i = \mathbf{a} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i,$$

where $\mathbf{a} = (a_1, \cdots, a_{12})'$, $\boldsymbol{\omega}_i = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3})'$, and

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

where $\boldsymbol{\xi} \overset{D}{=} N[\mathbf{0}, \boldsymbol{\Phi}]$, and

$$a_j = 0.5,\ \psi_j = 0.3,\ \text{for } j = 1, \cdots, 12$$

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

$$\phi_{ii} = 1.0,\ \phi_{ik} = 0.2,\ \text{for all } i,\ k,\ i \neq k;\ \psi_\delta = 0.3$$

Structural equation:

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

where

$$g(x) = (x/2)^3$$

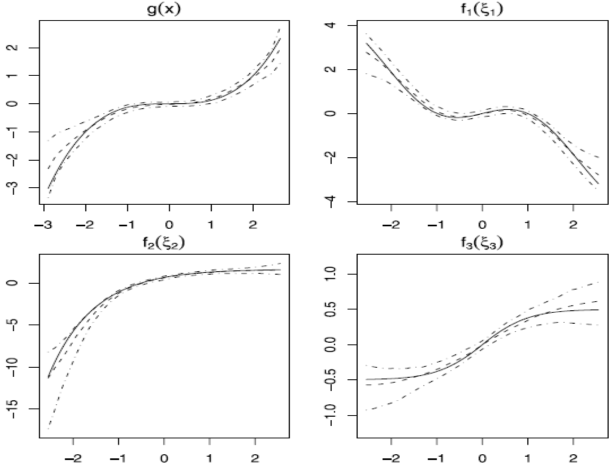
$$f_1(\xi) = sin(1.5\xi) - \xi$$

$$f_2(\xi) = 1.65 - \exp(\xi)$$

$$f_3(\xi) = -0.5 + \exp(2\xi)/[1 + \exp(2\xi)]$$

$n = 300$, $K = 20$, number of replications=100.

Par.	True	Bias	RMS	Par.	True	Bias	RMS
λ_{21}	0.8	-0.001	0.018	ϕ_{11}	1.0	-0.057	0.110
λ_{31}	0.8	-0.001	0.019	ϕ_{12}	0.2	-0.022	0.070
λ_{52}	0.8	0.022	0.044	ϕ_{13}	0.2	-0.010	0.064
λ_{62}	0.8	0.021	0.048	ϕ_{22}	1.0	-0.037	0.113
λ_{83}	0.8	0.022	0.050	ϕ_{23}	0.2	-0.003	0.058
λ_{93}	0.8	0.023	0.048	ϕ_{33}	1.0	-0.025	0.116
$\lambda_{11,4}$	0.8	0.007	0.049	ψ_{δ}	0.3	0.020	0.040
$\lambda_{12,4}$	0.8	0.016	0.046				



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!