Bayesian Analysis of Structural Equation Models with Nonlinear Covariates and Latent Variables

by

Xin-Yuan Song & Sik-Yum Lee

Department of Statistics, The Chinese University of Hong Kong

Acknowledgments: The work that is described in this paper was fully supported by a grant from the Research Grants Council of the HKSAR (Project No. CUHK 4243/02H).

Send correspondence to:

Prof. S.Y. Lee
Department of Statistics,
Chinese University of Hong Kong,
Shatin, N.T.,
Hong Kong.
E-mail: sylee@sparc2.sta.cuhk.edu.hk
Fax: 852-26037251
August, 2004
Abstract

In this article, we formulate a nonlinear structural equation model which can accommodate covariates in the measurement equation, and nonlinear terms of covariates and exogenous latent variables in the structural equation. The covariates can come from continuous or discrete distributions. A Bayesian approach is developed to analyze the proposed model. Markov chain Monte Carlo methods for obtaining Bayesian estimates and their standard error estimates, highest posterior density intervals, and a PP \( p \)-value are developed. The proposed methodology is illustrated by a real example. Detailed interpretation about the interaction terms is presented.

Keywords: Covariates, mixed continuous and dichotomous data, posterior simulation, interaction.
Introduction

Structural equation modeling is a multivariate method that allows one to investigate how the endogenous latent variables are related to or predicted by the exogenous latent variables, on the basis of non-experimental survey data. In past years, it is a widely appreciated tool and has drawn a great deal of attention in behavioral, social and psychological research, both in terms of theoretical developments and practical applications.

The most common and influential structural equation model (SEM) is perhaps the LISREL model and its program (Jöreskog & Sörbom, 1996). The appeal of LISREL is that it clearly defines two equations which play their unique role in formulating the substantive problem. The measurement equation explores and identifies the endogenous and exogenous latent variables, on the basis of the manifest variables. Then, the causal effects of the exogenous latent variables to the endogenous latent variables are investigated via the structural equation. Consequently, one of the main goals of SEM, namely relating or predicting the endogenous latent variables by the exogenous latent variables, can be achieved.

To establish better relationships or prediction, a useful extension of the basis SEM is the accommodation of covariates in the model. For the measurement equation that is used for exploring and/or identifying the latent variables, the inclusion of covariates is helpful in providing a clear relationship of the manifest and latent variables. As one of the main purpose of the structural equation is to achieve accurate prediction or relation among latent variables, the accommodation of covariates is even more important. Covariates can be explanatory variables such as age, gender, social status or other important variables that are defined by single manifest variables for relating or predicting the endogenous latent variables. These covariates may come from discrete and continuous distributions.

From a more general point of view, it is recently recognized that nonlinear terms of the exogenous latent variables are important in developing more correct and meaningful structural
equation for accessing the endogenous latent variables. For example, see Kenny and Judd (1984), Bagozzi, Baumgartner and Yi (1992), and articles in Schumacker and Marcoulides (1998), on the importance of including the interaction and quadratic effects. Hence analysis of nonlinear SEMs has received a lot of attention, including the methodological developments based on the productor indicator approach (see Jaccard & Wan, 1995; among others), the Bayesian approach (see for example, Lee & Song, 2003), and maximum likelihood (Lee & Zhu, 2002); as well as a comparative study of the above approaches via simulation studies (Lee, Song & Poon, 2004).

Clearly, in light of the motivation that is given above, it is useful to incorporate linear and nonlinear terms, not only among exogenous latent variables but also the covariates in the structural equation; especially for situations where some covariates that represent important predictors of the endogenous latent variables. In the article, we formulate a nonlinear SEM (NSEM) which includes a measurement equation that is defined with linear covariates and linear latent variables; and a structural equation that is defined by a general vector-valued function that subsumes general nonlinear terms of both the covariates and exogenous variables. In model sense, the proposed NSEM generalizes many existing NSEMs in the literature. For handling more complex data in practice, the methodology is developed in the context of mixed continuous, ordered categorical and dichotomous variables.

The methodological development is based on a Bayesian approach. The advantages of a Bayesian approach include allowing the use of genuine prior information in addition to that available in the observed data, and providing useful statistics, such as mean and percentiles, of the posterior distribution. In addition, as pointed out by many articles in Bayesian analysis of SEMs (see Dunson, 2000; Lee & Song, 2004; Schines, Hoijtink & Boomsma, 1999), the sampling-based Markov chain Monte Carlo (MCMC) methods do not rely on asymptotic theory, and hence give more reliable results for situations with small samples. Bayesian estimates of the unknown parameters are obtained from a sufficiently large number
of observations, which are sampled from the posterior distribution by the standard Gibbs sampler (Geman & Geman, 1984), and the Metropolis Hastings (MH) algorithm (Hastings, 1970; Metropolis, et al. 1953). In addition to the standard error estimates, the highest posterior density (HPD) intervals (see Casella & Berger, 1990) are also presented for assessing the variability of the Bayesian estimates. The goodness-of-fit of the posited model is assessed by the posterior predictive (PP) p-value that is developed by Gelman, Meng and Stern (1996).

The paper is organized as follows. The NSEM with a structural equation that accommodates nonlinear terms of covariates and exogenous latent variables is defined in the next section. Justification of the proposed formulation is discussed. Then, Bayesian methods for analyzing the NSEM is developed with conjugate prior distributions. Here, a hybrid algorithm that is based on the Gibbs sampler and the MH algorithm, and the PP p-value for goodness of fit, are derived. The section ‘An Illustrative Example’ presents a real example to illustrate the developed methodology, followed by a discussion in the last section. Some technical details are given in the appendices.

A NSEM with nonlinear covariates and latent variables

Similar to many LISREL type models, the proposed NSEM includes a measurement equation and a structural equation. For the measurement equation, we consider a random vector \( \mathbf{y}_i(p \times 1) \) of manifest variables that satisfies the following equation:

\[
\mathbf{y}_i = \mathbf{A} \mathbf{c}_i + \mathbf{\Lambda} \mathbf{\omega}_i + \mathbf{\epsilon}_i, \quad i = 1, \ldots, n
\]  

where \( \mathbf{c}_i(m_1 \times 1) \) is a vector of covariates, \( \mathbf{\omega}_i(q \times 1) \) is a vector of latent variables, \( \mathbf{\epsilon}_i(p \times 1) \) is a vector of error measurements, and \( \mathbf{A}(p \times m_1) \) and \( \mathbf{\Lambda}(p \times q) \) are unknown parameter matrices. It is assume that \( \mathbf{\epsilon}_i \) is distributed as \( \mathcal{N}(0, \mathbf{\Psi}_\epsilon) \), where \( \mathbf{\Psi}_\epsilon \) is a diagonal matrix with diagonal elements \( \psi_{\epsilon 1}, \ldots, \psi_{\epsilon p} \); and \( \mathbf{\epsilon}_i \) and \( \mathbf{\omega}_i \) are independent. The latent vector \( \mathbf{\omega}_i \) is partitioned into \( (\mathbf{\eta}_i^T, \mathbf{\xi}_i^T)^T \), where \( \mathbf{\eta}_i(q_1 \times 1) \) and \( \mathbf{\xi}_i(q_2 \times 1) \) are the vectors of the endogenous
and exogenous latent variables, respectively. To assess that possible important causal effects of a vector of covariates $x_i (m_2 \times 1)$ to $\eta_i$, the structural equation is defined by the following general equation:

$$\eta_i = B\eta_i + \Gamma F(x_i, \xi_i) + \delta_i, \quad i = 1, \cdots, n$$  \hspace{1cm} (2)

where $\delta_i (q_1 \times 1)$ is a vector of error measurements, $B (q_1 \times q_1)$ is a matrix of unknown parameters, $F(x_i, \xi_i) = (f_1(x_i, \xi_i), \cdots, f_r(x_i, \xi_i))^T$ is a vector-valued function with differentiable functions $f_1, \cdots, f_r$, and $\Gamma (q_1 \times r)$ is a matrix of unknown parameters. For simplicity, (2) can be expressed as

$$\eta_i = \Pi G(\eta_i, x_i, \xi_i) + \delta_i$$

where $\Pi = (B, \Gamma)$, and $G(y_i, x_i, \xi_i) = (\eta_i^T, F(x_i, \xi_i)^T)^T$. It is assumed that $\xi_i$ is distributed as $N[0, \Phi]$, and $\delta_i$ is distributed as $N[0, \Psi_\delta]$, where $\Psi_\delta$ is a diagonal matrix with diagonal elements $\psi_\delta_1, \cdots, \psi_{q_1}$, and $\delta_i$ and $\xi_i$ are independent. Similar to all other NSEMs (see Lee & Zhu, 2002; and Lee & Song, 2003), it is assumed that $B_0 = I - B_1$ is nonsingular and $|B_0|$ is independent of the elements of $B$. Examples of $B$ that are satisfied the above assumption are the upper or lower triangular matrices. Note that this conditional is assumed mainly for reducing the computational burden; and it can be relaxed by appropriate modifications.

The well-known LISREL model is clearly a special case of our proposed NSEM with no fixed covariates and $F(x_i, \xi_i) = \xi_i$. The factor analysis model with covariates considered by Sammel and Ryan (1996) only involved (1) without the important structural equation defined by (2). Clearly, the structural equation with the term $\Gamma F(x_i, \xi_i)$ is very general. A special case is

$$\eta_i = B\eta_i + \Gamma_1 x_i + \Gamma_2 F_2(\xi_i) + \delta_i,$$

which is equivalent to the structural equation given in Lee and Song (2003). A more concrete example of the general structural equation defined in (2) that is associated with $\eta_i = (\eta_i)$,
\( \xi_i = (\xi_{i1}, \xi_{i2})^T \), and \( x_i = (x_{i1}, x_{i2})^T \) is:

\[
\eta_i = \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_1 x_{i1} x_{i2} + \alpha_1 \xi_{i1} + \alpha_2 \xi_{i2} + \alpha_1 \xi_{i1} \xi_{i2} + \beta_1 x_{i1} \xi_{i1} + \beta_2 x_{i2} \xi_{i2} + \beta_3 x_{i1} \xi_{i1} \xi_{i2} + \beta_4 x_{i2} \xi_{i1} \xi_{i2} + \beta_5 x_{i1} x_{i2} \xi_{i1} \xi_{i2} + \delta_i
\]

Here, \( \Gamma = (\gamma_1, \gamma_2, \gamma_1, \alpha_1, \alpha_2, \alpha_1, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T \), and \( \mathbf{F}(x_i, \xi_i) = (x_{i1}, x_{i2}, x_{i1} x_{i2}, \xi_{i1}, \xi_{i2}, x_{i1} \xi_{i1}, x_{i2} \xi_{i2}, x_{i1} \xi_{i1} \xi_{i2}, x_{i2} \xi_{i1} \xi_{i2}, x_{i1} x_{i2} \xi_{i1} \xi_{i2})^T \). Of course, quadratic terms of elements \( x_i \) and \( \xi_i \) can be assessed via appropriately defined structural equations.

As the covariates \( x_i \) may come from any arbitrary distributions that give continuous data, ordered or unordered categorical data, the proposed NSEM can handle a wide range of situations. Under the context of nonlinear SEM, more care is needed to interpret the mean vector of \( y_i \), namely \( \mu \). Let \( A_k \) and \( \Lambda_k \) be the \( k \)th rows of \( A \) and \( \Lambda \), respectively. For \( k = 1, \ldots, p \), it follows from (1) that \( \mu_{ik} = \mathbb{E}(y_{ik}) = A_k c_i + \Lambda_k \mathbb{E}(\omega_i) \). Although \( \mathbb{E}(\xi_i) = 0 \), it follows from (2) that \( \mathbb{E}(\eta_i) \neq 0 \) if \( \mathbf{F}(x_i, \xi_i) \) is a nonlinear function of \( \xi_i \). Hence \( \mathbb{E}(\omega_i) \neq 0 \), and \( \mu_{ik} \neq A_k c_i \). Let \( \Lambda_k = (\Lambda_{kn}, \Lambda_{k\theta})^T \) be a partition of \( \Lambda_k \) that corresponds to the partition of \( \omega_i = (\eta_i^T, \xi_i^T)^T \). Because \( \mathbb{E}(\xi_i) = 0 \) and \( \eta_i = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{F}(x_i, \xi_i) \), it follows from (2) that

\[
\mu_{ik} = A_k c_i + \Lambda_{kn} \mathbb{E}(\eta_i) + \Lambda_{k\theta} \mathbb{E}(\xi_i) = A_k c_i + \Lambda_{kn}[(\mathbf{I} - \mathbf{B})^{-1} \Gamma] \mathbb{E}(\mathbf{F}(x_i, \xi_i)).
\]

(3)

As \( \mathbf{F}(x_i, \xi_i) \) is usually not very complicated in most practical applications, \( \mathbb{E}(\mathbf{F}(x_i, \xi_i)) \) is not very complex and the computation of \( \mu_{ik} \) is not difficult. See the illustrative example later.

One may consider the following indirect way for modeling covariates; by first augmenting \( y_i \) with \( x_i \), and then treating each element in \( x_i \) as an exogenous ‘latent variable’ that can be exactly measured with a single indicator:

\[
\begin{bmatrix}
  y_i \\
  x_i
\end{bmatrix} = \begin{bmatrix}
  A & 0 \\
  0 & 0
\end{bmatrix} c_i + \begin{bmatrix}
  \Lambda & 0 \\
  0 & I
\end{bmatrix} \begin{bmatrix}
  \omega_i \\
  x_i
\end{bmatrix} + \begin{bmatrix}
  \epsilon_i \\
  0
\end{bmatrix},
\]

(4)

and

\[
\eta_i = \mathbf{B} \eta_i + \Gamma^* \mathbf{F}^*(\xi^*_i) + \delta_i,
\]

(5)
where \( \xi^*_i = (\xi^T_i, x^T_i)^T \), \( F^* \) is a vector-valued differentiable functions, and \( \Gamma^* \) is a matrix of unknown constants. Then the model reduce to a special case of the NSEM given in Lee and Song (2003), and hence can be handled by their method. The approach may be workable for covariates that come from a normal distribution, which implies that \( \xi^*_i \) is normal. However, if \( x_i \) is non-normal, then \( \xi^*_i \) is non-normal. As the normality assumption of \( \xi^*_i \) is a crucial assumption of the model, the above approach cannot be applied to handle covariates in \( x_i \) that come from non-normal distributions. As a result, new methods for analyzing the nonlinear structural equation with non-normal covariates \( x_i \) are necessary. (As a minor separate point, even in the context of linear SEMs, it is rather difficult to reformulate the standard LISREL program for assessing non-normal covariates in the structural equation).

Based on similar reasoning given in Lee and Song (2003), we see that the proposed NSEM may be overparameterized under certain situations. For example, it does not allow intercepts to be existed in both the measurement and structural equations. Moreover, components of \( F(x_i, \xi_i) \) should be nonzero, and linearly independent. Moreover, appropriate elements in \( \Lambda \) and \( \Gamma \) should be fixed at appropriately known values for identification of the covariance structure.

To cope with the common data in substantive research, we allow that the exact measurements of some components in \( y \) are not available and the corresponding information is given by ordered categorical outcomes. In general, an ordered categorical variable \( z_h \) is defined with its underlying latent continuous random variable \( y_h \) by

\[
z_h = k \quad \text{if} \quad \alpha_k \leq y_h < \alpha_{k+1} \quad \text{for} \quad k = 0, \ldots, b_k - 1
\]

where the \( \alpha \)'s are unknown thresholds. A dichotomous outcome may be considered as a special case of an ordered categorical outcome with two categories and a fixed threshold at zero. Neither dichotomous variables nor ordered categorical variables are identified; however, well-known conditions, see Meng and Schilling (1996), and Song and Lee (2004a) among
others, can be applied to achieve identification of the model. For example, to identify a dichotomous variable \( z_h \) corresponding to its underlying continuous variable \( y_h \), we fixed \( \psi_{vh} = 1.0 \) according to the suggestion given in Song and Lee (2004a), or Song and Lee (2004c).

We note that in social, psychological, biological, and medical studies, correlated dichotomous data arise in many settings, ranging from measurements of random cross-section subjects to repeated measurements in longitudinal studies. Since the pioneer work of Ashford and Sowden (1970), numerous attempts have been made to solve the computational difficulty of evaluating the multivariate normal orthant probabilities that are involved with the multivariate dichotomous variables. For instance, methods that use much less restrictive covariance structures have been proposed to reduce the computational burden of evaluating the probabilities: see, for example, Kolakowski and Bock (1981). Moreover, Bock and Aitkin (1981) proposed the full-information item factor model by using exploratory factor analysis (EFA) model for the covariance structure. Recently, Bock and Gibbons (1996) and Gibbons and Wilcox-Gök (1998) extended the Bock and Aitkin (1981) model to a multivariate probit model that is equivalent to the following exploratory factor analysis model with fixed covariates: 
\[
y_i = A c_i + A \omega_i + \epsilon_i,
\]
in which \( y_i \) is unobservable but given by a vector of dichotomous variables \( z_i \), and \( \omega_i \) is distributed as \( N[0, I] \). Clearly, the above multivariate probit model is a special case of our NSEM.

Without lost of generality, suppose that \( y_i \) is equal to \((y_{0i}^T, y_{ui}^T)^T\), where \( y_{0i} \) is the subvector of \( y_i \) such that the original continuous measurements can be observed; whilst \( y_{ui} \) cannot be observed and its information is given by a vector of ordered categorical variables \( z_i \). Let \( D_0 = \{(y_{0i}, z_i), i = 1, \cdots, n\} \) be the observed-data set of mixed continuous and ordered categorical outcomes. The Bayesian approach for analysis of the proposed NSEM is developed in the next section, on the basis of this observed-data set.
Bayesian analysis of the model

One of our objective is to develop a Bayesian procedure to estimate the unknown parameter vector $\alpha$ which contains the unknown thresholds, and the unknown parameter vector $\theta$, which contains the parameters in $A, \Lambda, \Psi_\epsilon, \Phi$, and $\Psi_\delta$. In a Bayesian approach, the parameters in $\alpha$ and $\theta$ are considered to be random and it is possible to incorporate the prior information via the prior density of $(\alpha, \theta)$ in the posterior analysis. This is done by analyzing the following log posterior density of $\theta$ given $D_0$:

$$
\log p(\alpha, \theta | D_0) \propto \log p(D_0 | \theta, \alpha) + \log p(\theta, \alpha),
$$

where $p(D_0 | \theta, \alpha)$ is the observed-data likelihood function, and $p(\theta, \alpha)$ is the prior density of $\theta$ and $\alpha$. In this article, we follow the common approach to obtain the Bayesian estimate by computing the mean of the posterior distribution.

Due to the complexity of the model and the nature of the data, the posterior distribution is very complicated. It is very difficult and tedious to compute the posterior mean directly from $p(\alpha, \theta | D_0)$. Hence, the technique of data augmentation (Tanner & Wong, 1987) is employed in the posterior analysis. Let $\Omega = (\omega_1, \ldots, \omega_n)$ be the matrix of latent variables of the model, $Y_0 = (y_{01}, \ldots, y_{0m})$, $Z = (z_1, \ldots, z_n)$, and $Y_u = (y_{u1}, \ldots, y_{un})$ be the matrix of latent continuous measurements underlying the matrix of observed ordered-categorical data $Z$. In the analysis, the observed data $D_0 = (Y_0, Z)$ is augmented with $\Omega$ and $Y_u$, which can be considered as hypothetical missing data, to form a complete data set $(D_0, \Omega, Y_u)$. A large number of observations will be sampled from $p(\theta, \Omega, Y_u | D_0)$ by the Gibbs sampler (Geman & Geman, 1984) for obtaining the Bayesian estimates.

To implement the Gibbs sampler, we start with initial values $(\theta^{(0)}, \alpha^{(0)}, Y_u^{(0)}, \Omega^{(0)})$, simulate $(\theta^{(1)}, \alpha^{(1)}, Y_u^{(1)}, \Omega^{(1)}), \ldots$, and continue as follows. At the $j$-th iteration with current values $(\theta^{(j)}, \alpha^{(j)}, Y_u^{(j)}, \Omega^{(j)})$:

(a) Generate $\theta^{(j+1)}$ from $p(\theta | \alpha^{(j)}, Y_u^{(j)}, \Omega^{(j)}, D_0)$,
There are three main steps. But step (a) involves many components that correspond to \( \mathbf{A}, \Lambda, \Psi, \Pi, \Phi, \) and \( \Psi_\delta \). Hence it is divided into the following substeps which generate \( \Lambda^{(j+1)} \) from \( p(\Lambda, \Psi, \Pi, \Phi, \Psi_\delta) \) and \( \Psi_\delta \).

(b) Generate \((\alpha^{(j+1)}, Y_u^{(j+1)})\) from \( p(\alpha, Y_u, \theta^{(j+1)}, \Omega^{(j)}, D_0) \),

(c) Generate \( \Omega^{(j+1)} \) from \( p(\Omega | \theta^{(j+1)}, \alpha^{(j+1)}, Y_u^{(j+1)}, D_0) \).

At convergence, parallel sequences generated with different starting values should be mixed well. Another method to monitor convergence is described in Gelman (1996). Basically, the ‘estimated potential scale reduction (EPSR)’ values corresponding to the parameters are calculated sequentially as the runs proceed, and convergence is achieved when the EPSR values are less than 1.2. It is necessary to specify the prior distributions of components in \( \theta \) when deriving the conditional distribution of \( \theta \) given \( \alpha, Y_u, \Omega \) and \( D_0 \) in step (a). In general Bayesian analysis, the conjugate prior distributions have been found to be flexible and convenient (see, Broemeling, 1985). Conjugate prior distributions have been widely applied to many Bayesian analyzes in SEMs (Lee & Song, 2003; Song & Lee, 2004b). Hence, we shall use the following well-known conjugate prior distributions:

\[
\begin{align*}
p(\psi_{ek}^{-1}) & \overset{D}{=} Gamma[\alpha_{0ek}, \beta_{0ek}], \\
p(\Lambda_k | \psi_{ek}) & \overset{D}{=} N[A_{0k}, \psi_{ek}H_{0ek}], \\
p(A_k) & \overset{D}{=} N[A_{0k}, H_{0ak}], \\
p(\Phi) & \overset{D}{=} IW_q[R_0, \rho_0], \\
p(\psi_{bk}^{-1}) & \overset{D}{=} Gamma[\alpha_{0bk}, \beta_{0bk}], \\
p(\Pi_k | \psi_{sk}) & \overset{D}{=} N[\Pi_{0k}, \psi_{sk}H_{0sk}],
\end{align*}
\]

where ‘\( p(\cdot) \overset{D}{=} \)’ is defined as ‘the distribution of \( p(\cdot) \) is equal to’ , \( \psi_{ek} \) and \( \psi_{bk} \) are the \( k \)th
diagonal elements of $\Psi_\epsilon$ and $\Psi_\delta$ respectively, $A_k, \Lambda_k$ and $\Pi_k$ are $k$th rows of $A, \Lambda$ and $\Pi$, respectively; and $A_{0k}, \Lambda_{0k}, \Pi_{0k}, \alpha_{0k}, \beta_{0k}, \alpha_{0\delta k}, \beta_{0\delta k}, \Pi_{0k}, \rho_0$, and positive definite matrices $H_{0ak}, H_{0ek}, H_{0\delta k}$ and $R_0$ are known hyper-parameters whose values are given by the prior information. In general, prior information can be obtained from causal observation or theoretical consideration of experts, analysis of past data, or empirical information. By selecting appropriate $H_{0ak}, H_{0ek},$ or $H_{0\delta k}$, the magnitudes of the variances in the prior distributions that are associated with $A_k, \Lambda_k$ and $\Pi_k$ can be taken to be small when accurate prior information is available, or taken to be large when there is a lack of accurate prior knowledge.

For convenience, we follow the existing Bayesian analysis of SEMs with ordered categorical variables (see, for example, Song & Lee, 2004b, among others) to take an non-informative prior for the unknown thresholds. Hence, the prior distribution of $\alpha$ is proportional to a constant.

Based on the selected prior distributions, the full conditional distributions for implementing the Gibbs sampler are briefly derived in Appendix I. Note that with given $\Omega$ and $Y_u$, the NSEM as defined in (1) and (2) becomes the familiar regression model with continuous normal variables. Under some mild assumptions of prior distributions, it can be shown that conditional distributions of the components of $\theta$ are the familiar normal, Gamma, and inverted Wishart distributions. Simulating observations from the these distributions is fast and straightforward. However, the MH algorithm is required to implemented for the more complicated distribution that corresponds to $p(\alpha, Y_u|\theta, \Omega, D_0)$ and $p(\Omega|\theta, \alpha, Y_u, D_0)$.

Statistical inference of the model can be obtained on the basis of the simulated sample of observations from $p(\alpha, \theta, \Omega, Y_u|D_0)$, namely $\{(\theta^{(t)}, \Omega^{(t)}, \alpha^{(t)}, Y^{(t)}_u) : t = 1, \cdots, T\}$. The Bayesian estimate of $\theta$ as well as their numerical standard errors estimates can be obtained as

$$\hat{\theta} = \hat{E}(\theta|D_0) = T^{-1} \sum_{t=1}^{T} \theta^{(t)},$$

(8)
\[ \bar{\text{Var}}(\theta|D_0) = (T - 1)^{-1} \sum_{t=1}^{T} (\theta^{(t)} - \hat{\theta})(\theta^{(t)} - \hat{\theta})^T. \] 

(9)

Similarly, an estimate of \( \omega_i \) can be obtained from \( \{\Omega^{(t)} : t = 1, \cdots, T\} \). For revealing the variability of the Bayesian estimate, it is desirable to obtain the highest posterior density (HPD) interval of an individual parameter, see Cesella and Berger (1996). Achieving HPD intervals is not straightforward in the context of the current model. An algorithm that is based on the key idea of Chen, Shao and Ibrahim (2000) for constructing HPD is given in Appendix II.

A fundamental issue in analyzing SEMs is the assessment of the plausibility of a proposed model. The classical method in SEM associated with a non-Bayesian approach is to perform a goodness-of-fit test based on the asymptotic distribution of a test statistic that measures the discrepancy between the posited model and the sample covariance matrix. When dealing with more complicated models such as the current one, it is difficult to derive the asymptotic distribution of such a statistic. In Bayesian analysis, a simple method for assessing the goodness-of-fit of a posited model for fitting a data set is the posterior predictive p-values (PP p-values) developed by Gelman, Meng and Stern (1996), on the basis of the posterior assessment (Meng, 1994). They showed that this method is computationally and conceptually simple, and is useful in model-checking for a wide varieties of complicated situations. Moreover, the required computation is a by-product of the common Bayesian simulation procedures such as the Gibbs sampler. As suggested in Gelman et al. (1996), a PP p-value that is closed to 0.5 indicates very good fit of the posited model to the sample data. In practice, a posited model is acceptable if its corresponding PP p-value is in (0.3, 0.7). This statistic had been applied to Bayesian analyses of many SEMs (see, e.g. Lee & Song, 2004; Song & Lee, 2004b), and produced dependable results for assessing goodness-of-fit of the posited model. For completeness, the application of the PP p-value to our model is briefly outlined in Appendix III. In the illustrative example, this statistic is used to assess the goodness-of-fit
of the posited model.

An illustrative example

As an illustration of the Bayesian method for analyzing the proposed NSEM with fixed covariates, a portion of the data that were obtained from United Kingdom in the project WORLD VALUES SURVEY 1981-1984 and 1990-1993 was analyzed in our example. Seven variables related to respondents’ homelife, job satisfaction, and their attitude on job benefit and working environment were taken as manifest variables. As covariates are less important in the measurement equation, only the intercept $A = (a_1, \ldots, a_7)^T$ with all $c_i = 1.0$ is considered in defining (1). For the structural equation, a binary variable $x$ (with a shift of location according to the sample mean) that measures whether respondents can get comfort and strength from religion is incorporated as a covariate. Details of these variables are given in Appendix IV. For brevity, data points with missing entries were deleted, and the remaining sample size is 816. From the meaning of the corresponding questions, the first two manifest variables were indicators for the latent variable ‘home life’, the next two manifest variables were indicators for the latent variable ‘job satisfaction’, and the last three manifest variables were indicators for the latent variable ‘job benefit-environment attitude’. The manifest variables for ‘homelife’ and ‘job satisfaction’ were measured by a 10-point scale, and hence were treated as continuous. The remaining three manifest variables were dichotomous. For clear interpretation, we related these seven manifest variables with three latent variables via a $\Lambda$ with the following non-overlapping structure:

$$\Lambda^T = \begin{bmatrix} 1 & \lambda_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda_{42} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_{63} & \lambda_{73} \end{bmatrix},$$

where one’s and zero’s are fixed to identify the model. Parameters in $\Phi$ are $(\phi_{11}, \phi_{21}, \phi_{22})$. The covariance matrix $\Psi_\epsilon$ is taken to be a diagonal matrix, with diagonal elements $(\psi_{\epsilon 1}, \psi_{\epsilon 2}, \psi_{\epsilon 3})$. 
ψ_{1}, ψ_{3}, 1, 1, 1), in which ψ_{1}, ψ_{2}, ψ_{3} and ψ_{4} are unknown parameters, and the one’s are fixed to identify the dichotomous variables. As an illustration, we considered ‘homelife’ as the endogenous latent variable η, which is related with exogenous latent variables ‘job satisfaction, ξ_{1}’, ‘job benefit-environment attitude, ξ_{2}’, and the covariate about religion (x). The following nonlinear structural equation is used to assessed the interaction effects from the covariate x and the exogenous latent variables:

$$\eta_i = \gamma x_i + \alpha_1 \xi_{i1} + \alpha_2 \xi_{i2} + \alpha_{12} \xi_{i1} \xi_{i2} + \beta_1 x_i \xi_{i1} + \beta_2 x_i \xi_{i2} + \delta_i.$$  (10)

In this model, there are 25 unknown parameters. The path diagram describing the proposed NSEM is presented in Figure 1.

The Bayesian estimates of the parameters were obtained with conjugate prior distributions as presented in (7). The hyper-parameter values were produced by data-dependent prior inputs as in Lee and Song (2003). Note that although data-dependent priors have also been used in other Bayesian analyzes of statistical models (see e.g. Richardson & Green, 1997), we do not routinely recommend to use them for every practical problem. The convergence of the Gibbs sampler was monitored by the EPSR values (see Gelman (1996)). Figure 2 presents plots of these values against the iteration number. We observe that the sequence of observations converged after about 10000 iterations. To get the Bayesian estimates and the numerical standard errors, additional T = 10000 observations were collected after convergence, see (8) and (9). HPD intervals were also produced via implementation of the algorithm as described in Appendix II. The PP p-value is equal to 0.43, which indicates a good fit of the proposed model to the sample data.

The Bayesian estimates, the standard error estimates, and the HPD intervals are presented in Table 1. Some interesting interpretations are given below.

(i) All of the non-fixed factor loading estimates are quite large. From the HPD intervals, it can be seen that these factor loadings are different from zero. These results indicate
a strong association between each of the latent variables and their respective indicators. As expected, the standard errors of the loading estimates that correspond to the dichotomous manifest variables are relatively large, and the corresponding HPD intervals are relatively wide.

(ii) The estimated nonlinear structural equation is given by

\[ \eta_i = -0.197x_i + 0.621\xi_{i1} - 0.991\xi_{i2} + 0.603\xi_{i1}\xi_{i2} + 0.064x_i\xi_{i1} - 0.228x_i\xi_{i2}. \] (11)

From the standard error estimates and the HPD intervals, we observe that all the linear effects of the covariate and the exogenous latent variables, and the interaction effects corresponding to \( \xi_{i1}\xi_{i2} \) and \( x_i\xi_{i2} \) are quite different from zero, whilst the interaction effect corresponding to \( x_i\xi_{i1} \) is small. Hence, the interaction causal effect of religion \( (x_i) \) and ‘job satisfaction, \( \xi_{i1} \)’ is not significant when given the other linear and interaction effects in the structural equation. As \( \xi_{i2} \) is related to the dichotomous variables, the standard error estimate (=0.250) correspond to \( \hat{\beta}_2(= -0.228) \) is quite large, and the HPD interval, (-0.632, 0.175) is quite wide. However, as -0.228 is quite different from zero in magnitude, we tend to believe that the corresponding interaction effect is substantial.

(iii) Before giving more detailed interpretations of the nonlinear interaction terms, we note from the scales of the indicators in relation to “life, \( \eta \)” and “job satisfaction, \( \xi_1 \)” that a comparatively larger (positive) value of \( \eta \) or \( \xi_1 \) implies that the individual has better “life” or “job satisfaction”, respectively. From the scale of the indicators in relation to “job benefit-environment attitude, \( \xi_2 \)”, a comparatively smaller (negative) value of \( \xi_2 \) implies that the individual is more concerned about benefits and working environment. With the above understanding of the exogenous latent variables, it follows from \( \hat{\alpha}_1 = 0.621 \), and \( \hat{\alpha}_2 = -0.991 \) that job satisfaction and more concern about benefits and
working environment have a positive impact on “life, \( \eta \)”. From \( \hat{\alpha}_{12} = 0.603 \), “job satisfaction, \( \xi_1 \)” , and “job benefit-environment attitude, \( \xi_2 \)” have an interaction effect on “life, \( \eta \)” . The basic interpretation is that the ‘additive’ effect of the linear latent variables of \( \xi_1 \) and \( \xi_2 \) in the structural equation is inadequate to account for their relationships with the latent variable “life, \( \eta \)” , and an interaction term of \( \xi_1 \) and \( \xi_2 \) has to be added. Depending on various situations, this interaction term has different effects. For example, it is interesting to observe that for employees with good job satisfaction (positive \( \xi_{i1} \) ), and more concern about benefits and working environment (negative \( \xi_{i2} \) ) would have a too strong additive positive effect on “life, \( \eta \)” ; hence a negative adjustment via an interaction effect (0.603 times a positive \( \xi_{i1} \) and a negative \( \xi_{i2} \) would be negative) is necessary. For the covariate about religion, from the coding of the corresponding question, we note that an individual with a small (or negative) value of \( x_i \) gets more comfort and strength from religion, and vice versa. From \( \hat{\gamma} (= -0.197) \), we observe the expected finding that individuals who can gain comfort and strength from religion tends to have better life. The interpretation about the estimate of the interaction term (\( \hat{\beta}_2 = -0.228 \)) of \( x_i \) and \( \xi_{i2} \) is similar to the interpretation of \( \hat{\alpha}_{12} \).

(iv) From \( \hat{\phi}_{11} \), \( \hat{\phi}_{12} \) and \( \hat{\phi}_{22} \), the estimates of the elements in \( \Phi \), we note that the estimate of the correlation of ‘job satisfaction, \( \xi_1 \)” and ‘job benefit-environment attitude, \( \xi_2 \)” is 0.198. It indicates that job satisfaction and less concern about the benefit and environment are positively correlated.

(v) Now, we consider the interpretation of the intercepts and the means of the manifest variables. As an example, we consider \( \hat{\mu}_{i2} \). Since \( E(\xi_1) = E(\xi_2) = 0 \), \( \hat{a}_2 = 7.457 \), \( \hat{\lambda}_{21} = 0.768 \), \( \hat{\gamma} = -0.197 \), \( \hat{\alpha}_{12} = 0.603 \) and \( \hat{E}(\xi_{i1}\xi_{i2}) = \hat{\phi}_{12} = 0.186 \), we have

\[
\hat{\mu}_{i2} = \hat{a}_2 + \hat{\lambda}_{21}[\hat{\gamma}x_i + \hat{\alpha}_{12}\hat{E}(\xi_{i1}\xi_{i2})] = 7.920 - 0.151x_i
\]

This provides an estimate of the mean value of the second manifest variable of the
individual $i$, which depends on the value of the covariate $x_i$.

**Discussion**

In practice, substantive theory often involves latent variables that cannot be directly measured. Development and analysis of latent variable models for investigating the relationships of the observed variables and latent variables have received a great deal of past and recent attentions in statistics, behavioral and social-psychological research. The common approach in statistic focused on analyzing the contribution of the latent variables to the mean of the observed variables, as well as the effects of fixed covariates on the observed and/or latent variables. A primary interest is to estimate the regression coefficients that characterize the relationship among the variables and the fixed covariates. The normal theory random effects model (Laird & Ware, 1982) and the multivariate probit model (Bock & Gibbon, 1996; Gibbons & Wilcox-Göök, 1998) are the representative models of this approach. Another major approach is structural equation modeling (see, e.g. Bentler, 1983; Bollen, 1989; Jöreskog & Sörbom, 1996; among others) which focused on exploring hidden latent variables (factors) that cause the behaviors of observed variables, and investigating the relationships of latent variables among themselves and with the observed variables. The primary interests are to establish a meaningful model and to estimate the parameters in the underlying covariance structure. In this article, by formulating a nonlinear structural equation model that also accommodates covariates, we establish a general model that can be applied to analyze more subtle mean and covariance structures. Although our emphasis is on demonstrating the importance of the covariates, linear and nonlinear effects in explaining the endogenous latent variables in the structural equation, we have shown that our proposed model is a generalization of the full-information item factor model (Bock & Aitkin, 1981) and the multivariate probit model (Bock & Gibbons, 1996). By slight modification of the proposed model and the sampling-based method, methods for analyzing the normal theory mixed model can be
Based on the justifications that have been stated in the introduction, we develop a Bayesian approach to analyze the proposed model. However, the idea of data augmentation in augmenting the observed data with the latent quantities can be applied to develop a maximum likelihood approach with a Monte Carlo EM (MCEM) algorithm (Wei & Tanner, 1990). The Gibbs sampler and the related full conditional distributions that have been presented in this paper can be directly used to draw observations for approximating the more difficult E-step of the MCEM algorithm. The M-step can be completed via conditional maximization (Meng & Rubin 1994) or the standard Newton-Raphson algorithm.

The computational burden of the Bayesian sampling-based method is not as heavy as one may expect. For instance, using a SUN Enterprise, the computing time for getting the results of the illustrative example is 42 minutes. As the common software, such as the LISREL or the EQS program, cannot be applied to get the results, there are limitations for general users in applying the developed methodology in this article. However, the freely available software WinBUGS (Lunn, et al. 2000) is helpful; and our program is available upon request.

Mixed continuous and ordered categorical variables can be analyzed by the developed methodology. Due to the nonlinear latent variables, the distribution of the continuous variables is not normal. However, like almost all SEMs in the field, the conditional distribution of $y_i$ given the latent variables in $\omega_i$ is assumed to be normal. Extension of this conditional distribution to non-normal distribution, such as the exponential family distribution, to cope with the most general variables is an interesting topic for future research. Of course, development of useful methods that are more robust to the normal assumption is important.
References


Appendix I

Conditional Distribution of $\theta$

Conditional distribution of $p(\psi_{ek}, \Lambda_{k}|\cdot)$. For $k \neq h$, it is assumed that $(\psi_{ek}, \Lambda_{k})$ and $(\psi_{eh}, \Lambda_{h})$ are independent. Let $C = (c_1, \cdots, c_n)$, $U = (u_1, \cdots, u_n)$, $V = (v_1, \cdots, v_n)$, where $u_i = y_i - Ac_i$, $v_i = y_i - \Lambda c_i$, and $U_k$ and $V_k$ be the $k$-th row of $U$ and $V$, respectively. It can be shown by similar reasonings as Shi & Lee (1998), and Zhu & Lee (1998) that

\[ [A_k | Y, \Omega, A_k] \overset{D}{=} N[\mu_{ak}, \Sigma_{ak}], \quad [\Lambda_k | Y, \Omega, \psi_{ek}] \overset{D}{=} N[\mu_{ek}, \psi_{ek} \Sigma_{ek}], \]

where $\Sigma_{ak} = (H_{0ak} + CC^T)^{-1}$, $\mu_{ak} = \Sigma_{ak}[H_{0ak}A_{0k} + CV_k^T]$, $\Sigma_{ek} = (H_{0ek}^{-1} + \Omega\Omega^T)^{-1}$, $\mu_{ek} = \Sigma_{ek}[H_{0ek}^{-1}A_{0k} + \Omega U_k^T]$, and $\beta_{ek} = \beta_{0ek} + 2^{-1}(U_k^TU_k - \mu_{ek}^T\Sigma_{ek}^{-1}\mu_{ek} + \Lambda_{0k}^T H_{0ek}^{-1} A_{0k})$.

Conditional distribution of $p(\psi_{sk}, \Pi_{k}|\cdot)$, Let $\Omega_{(1)} = (\eta_1, \cdots, \eta_n)$, $\Omega_{(2)} = (\xi_1, \cdots, \xi_n)$, $\Omega_{(3)} = (G(y_1, x_1, \xi_1), \cdots, G(y_n, x_n, \xi_n))$, and $\Omega_k$ and $\Omega_{(1)k}$ be the $k$-th row of $\Omega$ and $\Omega_{(1)}$, respectively. For $k \neq h$, it is assumed that $(\psi_{sk}, \Pi_{k})$ and $(\psi_{sh}, \Pi_{h})$ are independent. Let $IW_{q_2}[, \cdot]$ denotes the inverted Wishart distribution with dimension $q_2$, it can be shown that

\[ [\Pi_k | \Omega, \psi_{sk}] \overset{D}{=} N[\mu_{sk}, \psi_{sk} \Sigma_{sk}], \quad [\psi_{sk}^{-1} | \Omega] \overset{D}{=} Gamma[n/2 + \alpha_{0sk}, \beta_{sk}]; \quad \text{for } k = 1, \cdots, q_1, \]

\[ [\Phi | \Omega] = [\Phi | \Omega_{(2)}] \overset{D}{=} IW_{q_2}((\Omega_{(2)} \Omega_{(2)}^T + R_0^{-1})n + \rho_0), \]

where $\Sigma_{sk} = (H_{0sk}^{-1} + \Omega_{(3)} \Omega_{(3)}^T)^{-1}$, $\mu_{sk} = \Sigma_{sk}[H_{0sk}^{-1}\Pi_{ok} + \Omega_{(3)} \Omega_{(1)k}]$, and $\beta_{sk} = \beta_{0sk} + 2^{-1}(\Omega_{(1)k} \Omega_{(1)k}^T - \mu_{sk}^T \Sigma_{sk}^{-1} \mu_{sk} + \Pi_{ok}^T H_{0sk}^{-1} \Pi_{ok})$.

Conditional Distribution of $(\alpha, Y_u)$

Let $Y_{uk}$ and $Z_k$ be the $k$-th row of $Y$ and $Z$, respectively; $\alpha_k$ is vector of unknown thresholds in $k$-th row. As $\Psi_e$ is diagonal, $(\alpha_k, Y_{uk})$ is independent of $(\alpha_h, Y_{uh})$ for $k \neq h$, and

\[ p(\alpha, Y_u|\cdot) = \prod_{k=1}^{s_1} p(\alpha_k, Y_{uk}|\cdot), \quad (A1) \]
where $Y_{uk} = \{y_{uik}, i = 1, \ldots, n_k\}$, in which $n_k$ is the number of $z_k$ in $Z_k$. Hence, $p(\alpha, Y_u \mid \cdot)$ can be obtained from (A1) and

$$p(\alpha_k, Y_{uk} \mid \cdot) \propto \prod_{i=1}^{n_k} \phi^* \left\{ \psi^{-1/2}_{zk} (y_{uik} - A_k^T c_i - \Lambda_k^T \omega_i) \right\} I_{\{\alpha_{si_k}, \alpha_{si_k+1}\}}(y_{uik}),$$

in which $I_A(y)$ is an index function which takes 1 if $y \in A$ and 0 otherwise, and $\phi^*(\cdot)$ is the density of $N[0, 1]$. 

**Conditional distribution** $p(\Omega \mid \theta, \alpha, Y_u, D_0)$. 

It is noted that given $Y_u$, the underlying model becomes one with continuous data that is relatively easy to handle. Moreover, as $\omega_i, i = 1, \ldots, n$ are mutually independent, and they are not depending on $\alpha$, we have

$$p(\Omega \mid \cdot) = \prod_{i=1}^{n} p(\omega_i \mid \theta, y_i) \propto \prod_{i=1}^{n} p(y_i \mid \theta, \omega_i) \cdot p(\eta_i \mid \xi_i, \theta) \cdot p(\xi_i \mid \theta).$$

It follows that $p(\omega_i \mid \cdot)$ is proportional to

$$\exp \left[ -\frac{1}{2} \left( \xi_i^T \Phi^{-1} \xi_i + (y_i - Ac_i - \Lambda \omega_i)^T \Psi^{-1} (y_i - Ac_i - \Lambda \omega_i) ight. 
\left. + [\eta_i - B \eta_i - \Gamma F(x_i, \xi_i)]^T \Psi_{\delta}^{-1} [\eta_i - B \eta_i - \Gamma F(x_i, \xi_i)] \right] \right].$$

22
Appendix II

Let $\nu$ be a component in $\theta$, the $(1 - \alpha)$ HPD interval for $\nu$ is given by $H(\alpha) = \{ \nu : p(\nu|D_0) \geq c_0 \}$, where $c_0$ is chosen so that

$$1 - \alpha = \int_{H(\alpha)} p(\nu|D_0) d\nu.$$ 

In this paper, HPD intervals presented in the real example were estimated via the Chen-Shao algorithm (Chen & Shao, 1999) as follows:

Step 1. Given an MCMC sample $\{\nu_i, i = 1, \cdots, T\}$ from $p(\nu|D_0)$ obtained via the proposed Gibbs sampler-MH algorithm.

Step 2. Sort $\{\nu_i, i = 1, \cdots, T\}$ to obtain the ordered values: $\nu_{(1)} \leq \nu_{(2)} \leq \cdots \leq \nu_{(T)}$.

Step 3. Compute the $100(1 - \alpha)\%$ credible intervals $(\nu_{(j)}, \nu_{(j+[(1-\alpha)T])})$, for $j = 1, 2, \cdots, T - [(1 - \alpha)T]$.

Step 4. The $100(1 - \alpha)\%$ HPD interval is the one with the smallest interval width among all credible intervals.
Appendix III

Meng (1994) introduced a Bayesian counterpart of the classical $p$-value by defining a posterior predictive $p$-value that depends both on the data and the unknown parameters. Suppose that $H_0$ is the null hypothesis that the proposed model defined in (1) and (2) is plausible, the posterior predictive $p$-value is defined as

$$p_B = P\{D(Y^{rep} | \Omega, Y_u, \theta) \geq D(Y | \Omega, Y_u, \theta) | D_0, H_0)\},$$

where $Y^{rep}$ denotes a replication of $Y$ and $D(\cdot | \cdot)$ is a discrepancy variable. For our model, we choose the following $\chi^2$ discrepancy variable

$$D(Y^{rep} | \Omega, Y_u, \theta) = \sum_{i=1}^{n} (y_i^{rep} - Ac_i - \Lambda \omega_i)')^T \Psi_{e}^{-1}(y_i^{rep} - Ac_i - \Lambda \omega_i),$$

which is distributed as a $\chi^2(pn)$ distribution. The posterior predictive (PP) $p$-value based on this discrepancy variable is given by

$$p_B(D_0) = \int P[\chi^2(pn) \geq D(Y | \Omega, Y_u, \theta)] p(\Omega, Y_u, \theta | D_0) d\Omega dY_u d\theta.$$

A Rao-Blackwellized type estimate of the PP $p$-value is equal to

$$\hat{p}_B(D_0) = T^{-1} \sum_{t=1}^{T} P[\chi^2(pn) \geq D(Y_0, Y_u^{(t)} | \Omega^{(t)}, \theta^{(t)})]. \tag{12}$$

The computation of $\hat{p}_B(D_0)$ is straightforward, since $D(Y_0, Y_u^{(t)} | \Omega^{(t)}, \theta^{(t)})$ can be calculated in each iteration and the tail-area probability of $\chi^2$ distribution can be obtained in any standard statistical software. Refer to Meng (1994), Gelman, et al. (1996) for more details about this general procedure.
Appendix IV

Seven manifest variables and one fixed covariate in the ICPSR data set: The number of the variable corresponding to the original data set is given in parenthesis at the end of each statement.

\( y_1 \): Overall, how satisfied or dissatisfied are you with your home life? (V180)

[ Dissatisfied — 1 2 3 4 5 6 7 8 9 10 — Satisfied ]

\( y_2 \): All things considered, how satisfied are you with your life as a whole these day? (V96)

[ Dissatisfied — 1 2 3 4 5 6 7 8 9 10 — Satisfied ]

\( y_3 \): Overall, how satisfied or dissatisfied are you with your job? (V116)

[ Dissatisfied — 1 2 3 4 5 6 7 8 9 10 — Satisfied ]

\( y_4 \): How free are you to make decisions in your job? (V117)

[ None at all — 1 2 3 4 5 6 7 8 9 10 — A great deal ]

Here are some aspects of a job that people say are important. Which ones you personally think are important in a job?

\( y_5 \): Good pay. (V99) [ Mentioned 0 / 1 Not mentioned ]

\( y_6 \): Pleasant people to work with. (V100) [ Mentioned 0 / 1 Not mentioned ]

\( y_7 \): Not too much pressure. (V101) [ Mentioned 0 / 1 Not mentioned ]

\( x \): Do you find that you get comfort and strength from religion? (V177) [ Yes 0 / 1 No ]
Table 1: Bayesian estimates (EST), the standard error estimates (SE) and HPD intervals.

<table>
<thead>
<tr>
<th>Par</th>
<th>EST</th>
<th>SE</th>
<th>HPD</th>
<th>Par</th>
<th>EST</th>
<th>SE</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>8.120</td>
<td>0.06</td>
<td>(8.02, 8.22)</td>
<td>$\gamma$</td>
<td>-0.197</td>
<td>0.10</td>
<td>(-0.37, -0.03)</td>
</tr>
<tr>
<td>$a_2$</td>
<td>7.457</td>
<td>0.06</td>
<td>(7.36, 7.55)</td>
<td>$\alpha_1$</td>
<td>0.621</td>
<td>0.09</td>
<td>(0.48, 0.76)</td>
</tr>
<tr>
<td>$a_3$</td>
<td>7.425</td>
<td>0.07</td>
<td>(7.31, 7.54)</td>
<td>$\alpha_2$</td>
<td>-0.991</td>
<td>0.27</td>
<td>(-1.43, -0.49)</td>
</tr>
<tr>
<td>$a_4$</td>
<td>6.972</td>
<td>0.09</td>
<td>(6.83, 7.10)</td>
<td>$\alpha_{12}$</td>
<td>0.064</td>
<td>0.08</td>
<td>(-0.08, 0.19)</td>
</tr>
<tr>
<td>$a_5$</td>
<td>-0.755</td>
<td>0.06</td>
<td>(-0.84, -0.70)</td>
<td>$\beta_1$</td>
<td>-0.228</td>
<td>0.25</td>
<td>(-0.63, 0.17)</td>
</tr>
<tr>
<td>$a_6$</td>
<td>-0.462</td>
<td>0.05</td>
<td>(-0.55, -0.38)</td>
<td>$\beta_2$</td>
<td>0.603</td>
<td>0.15</td>
<td>(0.36, 0.85)</td>
</tr>
<tr>
<td>$a_7$</td>
<td>0.989</td>
<td>0.07</td>
<td>(0.88, 1.12)</td>
<td>$\lambda_{21}$</td>
<td>0.908</td>
<td>0.08</td>
<td>(0.78, 1.05)</td>
</tr>
<tr>
<td>$\lambda_{42}$</td>
<td>0.768</td>
<td>0.09</td>
<td>(0.64, 0.91)</td>
<td>$\phi_{11}$</td>
<td>2.489</td>
<td>0.06</td>
<td>(1.95, 2.97)</td>
</tr>
<tr>
<td>$\lambda_{63}$</td>
<td>0.644</td>
<td>0.17</td>
<td>(0.37, 0.93)</td>
<td>$\phi_{12}$</td>
<td>0.186</td>
<td>0.06</td>
<td>(0.07, 0.30)</td>
</tr>
<tr>
<td>$\lambda_{73}$</td>
<td>0.894</td>
<td>0.22</td>
<td>(0.56, 1.28)</td>
<td>$\phi_{22}$</td>
<td>0.355</td>
<td>0.07</td>
<td>(0.23, 0.47)</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>1.120</td>
<td>0.18</td>
<td>(0.81, 1.40)</td>
<td>$\psi_{5}$</td>
<td>0.718</td>
<td>0.19</td>
<td>(0.43, 1.01)</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>1.391</td>
<td>0.15</td>
<td>(1.16, 1.64)</td>
<td>$\psi_{6}$</td>
<td>0.718</td>
<td>0.19</td>
<td>(0.43, 1.01)</td>
</tr>
<tr>
<td>$\psi_{13}$</td>
<td>1.625</td>
<td>0.26</td>
<td>(1.19, 2.04)</td>
<td>$\psi_{14}$</td>
<td>4.869</td>
<td>0.29</td>
<td>(4.39, 5.32)</td>
</tr>
</tbody>
</table>
Figure 1: Path diagram of the SEM with interactions of covariates and latent variables
Figure 2: EPSR values against the number of iterations in the analysis of ICPSR data.