Analysis of Generalized Nonlinear Structural Equation Models by Using Bayesian Approach with Application

Thanoon Y. Thanoon
Northern Technical University, Technical College of Management, Mosul, Iraq
thanoon.younis80@gmail.com

Robiah Adnan
Department of Mathematical Science, Faculty of Science
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
robiaha@utm.my

Abstract

In this paper, Bayesian analysis is used in nonlinear structural equation models with two population of data and the Gibbs sampling method is applied for estimation and model comparison. Hidden continuous normal distribution (censored normal distribution) is used to solve the problem of ordered categorical data in Bayesian multiple group SEMs and compared with the method that treats ordered categorical variables as a continuous normal distribution. Statistical inferences, which involve the estimation of parameters and their standard errors, and residuals analyses for testing the posited model are discussed. The proposed procedure is illustrated using real data with the results obtained from the WinBUGS program.

Keywords: Structural Equation Models, Bayesian Analysis, Censored Normal Distribution, Gibbs Sampling, Ordered Categorical Variables.

1. Introduction

Structural equation models (SEMs) (Bollen and Paxton, 1998; Lee, 2007) constitute a statistical methodology for modelling multivariate correlated data to assess the interrelationships among observed and latent variables.

At present, most statistical theory and computer software in the field of SEMs are based on models that involve nonlinear relationships among the manifest and the latent variables. More statistically sound methods for nonlinear SEMs and factor analysis have been proposed by Lee & Tang, (2006), Lee & Tang, (2007), Cai *et al.*, (2008), Lee *et al.*,(2010), Thanoon & Adnan, (2015), Thanoon & Adnan, (2016a), Thanoon & Adnan, (2016b), Thanoon & Adnan, (2016c), Thanoon & Adnan, (2016d), Thanoon *et al.*, (2017).

Lee & Song, (2002) proposed a method using Bayesian approach for multi-group nonlinear factor analysis. A Bayesian approach for a general multi-group nonlinear factor analysis model is developed. Joint Bayesian estimates of the factor scores and the structural parameters subjected to some constraints across different groups are obtained simultaneously. A hybrid algorithm that merges the Metropolis- Hastings algorithm and the Gibbs sampler is conducted to produce these joint Bayesian estimates. It is shown that this algorithm is computationally efficient. The Bayes factor approach is introduced for comparing models under various degrees of invariance across groups. The Schwarz criterion (BIC), a simple and useful approximation of the Bayes factor, is calculated on

the basis of simulated observations from the Gibbs sampler. Efficiency and flexibility of the proposed Bayesian procedure are illustrated by some simulation results and a real-life example. Song and Lee, (2006) discussed multiple group nonlinear structural equation models with missing continuous and dichotomous data that involve data that are missing at random using maximum likelihood approach, as well as, he demonstrated the newly developed methods for estimation and model comparison by a simulation study and a real data application. Lee, (2007) used underlying latent continuous normal distribution with truncation to solve the problem of ordered categorical variables in Bayesian multiple group nonlinear structural equation models, as well as, Gibbs sampling method is used to estimate the parameters. Lu *et al.,* (2012) used Bayesian analysis of multiple group nonlinear structural equation models with application to behavioral finance and treated the ordered categorical variables as a continuous normal distribution. The proposed method is used to investigate the relationships among all identified influential factors that have an impact on the motivation for insider trading within the framework of behavioural finance (Thanoon and Adnan, 2015).

For the structural equation model with ordered categorical data, an important generalization is to extend the model to permit analysis of multiple groups or populations of individuals simultaneously. Multiple group analysis is important in various applications, such as cross-cultural research. It is very interesting to investigate whether the measurement items (which are often on Likert scales of a categorical nature) are of different cultures. To achieve this goal requires statistical methods for testing various hypotheses in the multi-sample structural equation models with ordered categorical variables.

The Bayesian approach is developed with the Gibbs sampler (Geman & Geman, 1984) algorithm, in which the continuous latent measurements and the latent variables in different groups are treated as hypothetical missing data. Non-informative priors are used for the thresholds and conjugate priors are used for the structural parameters.

Theoretically, the importance of generalizing nonlinear structural equation models to nonlinear models that include nonlinear terms of the latent variables is obvious. Practically, nonlinear relationships, such as quadratic and interaction terms, among the variables are important in establishing the substantive theory in many areas. The rapid growth of SEMs is due to the demand of subtle models and the related statistical methods for solving complex research problems in various fields.

A major breakthrough for posterior simulation is the idea of data augmentation proposed by Tanner and Wong (1987). The strategy is to treat latent quantities as hypothetical missing data and to augment the observed data with them so that the posterior distribution based on the complete data set is relatively easy to analyze. The feature that makes SEMs different from the common regression model and the simultaneous equation model is the existence of random latent variables.

The data augmentation provides a useful approach to cope with the problem that is induced by latent variables. By augmenting the observed variables in complicated SEMs with the latent variables that are treated as hypothetical missing data, we can obtain the Bayesian solution based on the complete data set. Theoretically, the importance of

generalizing nonlinear structural equation models to nonlinear models that include nonlinear terms of the latent variables is obvious. Practically, nonlinear relationships, such as quadratic and interaction terms, among the variables are important in establishing the substantive theory in many areas. The rapid growth of SEMs is due to the demand of subtle models and the related statistical methods for solving complex research problems in various fields.

The main objective of this paper is to propose a Bayesian approach for analysing multiple group nonlinear SEMs with ordered categorical variables. The Deviance Information Criterion (DIC; see Spiegelhalter *et al.*, 2002) will be used for model comparison.

The main idea in handling the ordered categorical variables in the Bayesian analysis is to treat the underlying latent continuous measurements as hypothetical missing data and augment them with the observed data in the posterior analysis.

The paper is organized as follows. The Model Description is described in Section 2. The Bayesian estimation of multiple group structural equation models that contain nonlinear models is described in Section 3. The normal distribution is presented in section 4. The truncated normal distribution is explained in Section 5. The censored normal distribution is presented in Section 6. The comparison of models using DIC is described in Section 7. A real example study is presented in Section 8. Analysis of a real data are discussed in Section 9. The results and discussion are described in Section 10, and some concluding remarks are given in Section 11.

2. Model Description

Consider a set of G populations which may be different nations, states or regions, cultural or socio-economic groups, groups receiving different treatments, etc. The definition of the multiple group structural equation models is given by:

For $g = 1, 2, ..., G$, and $i = 1, 2, ..., N_g$, let $y_i^{(g)}$ be the $p \times 1$ random vector of observed variables that relate to the ith observation in the gth group. For each $g = 1, 2, ..., G$, $y_i^{(g)}$ is related to latent variables in a $q \times 1$ random vector $w_i^{(g)}$ by the next measurement equation: (Lee & Song, 2012; Lee, 2007; Lee *et al.*, 2010)

$$y_i^{(g)} = \mu^{(g)} + \Lambda^{(g)} \omega_i^{(g)} + \varepsilon_i^{(g)}, \qquad i = 1, ..., n \qquad (1)$$

Where $u^{(g)}$ is the intercepts vector, $\Lambda^{(g)}$ is the parameter matrix of regression coefficients that reflect the relation of manifest variables in $y_i^{(g)}$ with the latent variables in $w_i^{(g)}$, and $\varepsilon_i^{(g)}$ is a random vector of the measurement errors. It is presumed that $w_i^{(g)}$ and $\varepsilon_i^{(g)}$ are independent while the distribution of $\varepsilon_i^{(g)}$ is $N[0, \Psi_\varepsilon^{(g)}]$, where $\Psi_\varepsilon^{(g)}$ is a diagonal covariance matrix. Let $w_i^{(g)} = (\eta_i^{(g)}, \xi_i^{(g)T})^T$, where $\eta_i^{(g)}$ stand for the $q_1 \times 1$ vector of endogenous latent variables and $\xi_i^{(g)}$ stand for the $q_2 \times 1$ vector of the exogenous latent variables. Note that it is naturally presumed that q_1 and q_2 are independent of g, that is, they are the same for each group. So, to assess the effects of the nonlinearity of latent

variables in $\eta_i^{(g)}$ to $\xi_i^{(g)}$ we consider a nonlinear SEM with the following nonlinear structural equation:

$$\eta_i^{(g)} = \Pi^{(g)}\eta_i^{(g)} + \Gamma^{(g)}H(\xi_i^{(g)}) + \delta_i^{(g)} = \Lambda_w^{(g)}H^*(w_i^{(g)}) + \delta_i^{(g)} \tag{2}$$

Where $\Pi^{(g)}$ and $\Gamma^{(g)}$ are unknown parameter matrices, and $\delta_i^{(g)}$ is the random vector of error measurements that is independent of $\xi_i^{(g)}$; $\Lambda_w^{(g)} = (\Pi^{(g)}, \Gamma^{(g)})$,

$H^*(w_i^{(g)}) = (\eta_i^{(g)T}, H(\xi_i^{(g)})^T)^T$ and $H(\xi_i^{(g)}) = (h_1(\xi_i^{(g)}), ..., h_a(\xi_i^{(g)}))^T$, in which the *hs* are nonzero and known differentiable functions that include polynomials.

It is expected that the vector-valued functions $\boldsymbol{H}$ and $\boldsymbol{H}^*$ do not depend on g. However, different groups can require different linear or nonlinear function of $\xi_i^{(g)}$ by major appropriate of $\boldsymbol{H}$ or $\boldsymbol{H}^*$ and assigning zero values to appropriate elements in $\Gamma^{(g)}$. For instance, let $w_i^{(g)} = (\eta_i^{(g)}, \xi_{i1}^{(g)}, \xi_{i2}^{(g)})^T$ for $g = 1, 2$. Suppose that the structural equations in the first and second groups are given as:

$$\eta_i^{(1)} = \gamma_1^{(1)}\xi_{i1}^{(1)} + \gamma_2^{(1)}\xi_{i2}^{(1)} + \gamma_3^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)} + \delta_i^{(1)} \tag{3}$$

$$\eta_i^{(2)} = \gamma_1^{(2)}\xi_{i1}^{(2)} + \gamma_2^{(2)}\xi_{i2}^{(2)} + \gamma_4^{(2)}\xi_{i1}^{(2)}\xi_{i1}^{(2)} + \delta_i^{(2)} \tag{4}$$

To define these structural equations, we consider

$H(\xi_i^{(g)}) = (\xi_{i1}^{(g)}, \xi_{i2}^{(g)}, \xi_{i1}^{(g)}\xi_{i2}^{(g)}, \xi_{i1}^{(g)}\xi_{i1}^{(g)})^T$, $\Gamma^{(1)} = (\gamma_1^{(1)}, \gamma_2^{(1)}, \gamma_3^{(1)}, 0)$ and $\Gamma^{(2)} = (\gamma_1^{(2)}, \gamma_2^{(2)}, 0, \gamma_4^{(2)})$.

In reality, the models in different groups typically have similar linear and nonlinear of $\xi_i^{(g)}$. For handling really complicated different nonlinear functions in the models, this assumption can be relaxed with minor changes. Likewise, as for the models, it is assumed that $\boldsymbol{I} - \Pi^{(g)}$ is a non-singular matrix where the determinant is not dependent on elements in $\Pi^{(g)}$. It is further assumed that $\xi_i^{(g)}$ is distributed as $N[\boldsymbol{0}, \Phi^{(g)}]$ and $\delta_i^{(g)}$ is distributed $N[\boldsymbol{0}, \Psi_\delta^{(g)}]$, where $\Psi_\delta^{(g)}$ is a diagonal covariance matrix.

To handle the ordered categorical outcomes, suppose that $y_i^{(g)}$ is an $s \times 1$ subvector of unobservable continuous responses, the information of which is reflected by an observable ordered categorical vector $z_i^{(g)}$. In a generic sense, an ordered categorical variable $z_m^{(g)}$ is defined with its underlying latent continuous random variable $y_m^{(g)}$ by:

$$z^{(g)} = \begin{bmatrix} z_1^{(g)} \\ \vdots \\ z_s^{(g)} \end{bmatrix} \quad if \quad \begin{bmatrix} \alpha_{1,z_1^{(g)}}^{(g)} < y_1^{(g)} \le \alpha_{1,z_1^{(g)}+1}^{(g)} \\ \vdots \\ \alpha_{s,z_s^{(g)}}^{(g)} < y_s^{(g)} \le \alpha_{s,z_s^{(g)}+1}^{(g)} \end{bmatrix} \tag{5}$$

where $\{-\infty = \alpha_{m,1}^{(g)} < \alpha_{m,2}^{(g)} < ... < \alpha_{m,b_m}^{(g)} < \alpha_{m,b_m+1}^{(g)} = \infty\}$ is the set of threshold parameters that define the categories, and b_m is the number of categories for the ordered categorical variable $z_m^{(g)}$. Note that for each ordered categorical variable, the number of thresholds is the same for each group. However, the categories can be unequally spaced under this formulation.

It has been pointed out by (Lee *et al*., 1990) that single-sample models with ordered categorical variables are not identified without imposing identification conditions. This is also the case for multi-sample models. To solve this problem, we use the common method (see, for example, Lee *et al.*, 1995; Shi & Lee, 1998) of fixing some thresholds at pre-assigned values. For convenience, it is assumed that the positions of the fixed elements are the same for each group.

Multiple Group data come from a comparatively smaller number of groups of populations. The number of observations within each group is normally large and assumed independent. The primary purpose of the multiple group data analysis is to investigate the similarities or differences among the models in the different groups. As a consequence, the statistical inferences emphasized in analyzing multiple group of the SEMs are different from those in analyzing two-level SEMs. Analysis of the multiple group is a major issue in structural equation modeling because it is useful for investigating the behaviors of different groups for example of employees, cultures, treatment groups, and so on. Though, the testing of hypotheses about the different invariances among the models with different groups is the focus point.

This issue can be described as a model comparison problem, and also effectively addressed by the Bayes factor or DIC in a Bayesian approach. The benefit of the Bayesian model comparison over the Bayes factor or DIC is that the hypotheses of nonnested models can be compared. Hence, it is not necessary to follow hypotheses hierarchy to assess the invariance for the SEMs in different groups (Song & Lee, 2012; Wang & Wang, 2012).

3. Bayesian Analysis of Multiple Group Nonlinear Structural Equation Models

Allow $Z = (z_1,...,z_n)$ stand for the matrix of ordered categorical data. Then, allow $Y = (y_1,...,y_n)$ and $\Omega = (w_1,...,w_n)$ be the latent continuous measurements and latent variables matrices. Such that, the observed data, as represented by [Z] is improved according to the latent data $[Y, \Omega]$ through the posterior analysis. Thus in order to define the Bayesian approach for the proposed SEM, allow θ to act as the vector containing unknown parameters. In explanation, the Bayesian approach, would define $p(\theta)$ such that θ is considered a random with prior distribution and prior density function. Thus, the the related assumptions can be based on the observed data for Z and $p(\theta)$. So, allow Let $p(Z,\theta)$ represent the joint probability density function of both Z and θ with reference to different M$_k$. Based on a well-known identity in probability,

$$\log p(\theta|Z) \propto \log p(Z|\theta) + \log p(\theta) \tag{6}$$

This yields the posterior density function $p(\theta|Z)$, or unknown parameters. It also employs the use of sample information and prior knowledge, via the likelihood function $p(Z|\theta)$, and the prior density function $p(\theta)$. It should be noted, however, that $p(Z|\theta)$ is dependent on sample size, where $p(\theta)$ is not. As such, for problems with a large sample, $p(\theta)$ is less significant and $p(\theta|Z)$, the posterior density function, is more relevant, as it is most similar to the likelihood function $p(Z|\theta)$. So, both the Bayesian approach and ML model are asymptotically equivalent, and thus contain the same optimal asymptotical properties. However, continue to note that $p(\theta)$ is significant with regard to the Bayesian approach when the sample size is reduced or when the information derived from Z is ordered categorical data.

In this paper, MCMC methods are applied by letting y_i act as an unobserved variable, which corresponds with the manifest ordered categorical variables, as they are found in z_i. Also, one must allow $Y = (y_1,...,y_n)$ and $\Omega = (w_1,...,w_n)$. When drawing a sufficient, and generally large, body of observations If we can draw a sufficiently large number of observations, represented by $\{(\theta^{(t)},\Omega^{(t)},Y^{(t)}); t = 1,...,T\}$, from the joint posterior distribution, defined by $p(\theta,\Omega,Y|Z)$, then the Bayesian estimate for θ as well as any standard error estimates can be derived from the sample mean and variance matrices, respectively.

$$\widehat{\theta} = T^{-1}\sum_{t=1}^{T}\theta^{(t)}, \quad \text{var}(\theta|Z) = (T-1)^{-1}\sum_{t=1}^{T}(\theta^{(t)}-\widehat{\theta})(\theta^{(t)}-\widehat{\theta})'. \tag{7}$$

This means that it is necessary to specifically identify the prior distribution for the related components in θ, even if developing the conditional distribution, $(\theta|\Omega,Y,Z)$. In generally, during Bayesian analysis, the conjugate prior distributions have proven to be both malleable and suitable to the purpose (Broemeling, 1985). This kind of prior distribution has been widely applied to many Bayesian analysis in structural equation models, (see (Lee and Song, 2004; Song and Lee, 2007)).

Hence, the following well-known conjugate prior distributions are used:

$$p(\mu) \square N[\mu_0, H_{0\mu}], \quad p(\lambda_k) \square N[\lambda_{0k}, H_{0k}], \quad p(\lambda_{\xi k}|\psi_{\delta k}) \square N[\lambda_{0\xi k}, \psi_{\delta k} H_{0\xi k}],$$

$$p(\Phi^{-1}) \square W_q[R_0, \rho_0], \quad p(\psi_{\delta k}^{-1}) \square Gamma[\alpha_{0k}, \beta_{0k}] \tag{8}$$

Given the definition that $p(\cdot)\square$ is the probability of $p(\cdot)$, and that $p(\cdot)$ is distributed according to, $\psi_{\delta k}$, which is the *k*th diagonal element of Ψ_δ, $\lambda_k{}'$ and, given that $\lambda_{\xi k}{}'$ are the *k*th rows of Λ and Λ_ξ respectively. The following can be derived:

$H_{0\mu} = diag(\sigma_{01}^2,...,\sigma_{0p}^2)$, and $\mu_0, \lambda_{0k}, \lambda_{0\xi k}, \alpha_{0k}, \beta_{0k}, \rho_0, \sigma_{0k}, H_0, H_{0\xi k}$, and R_0 are assumed to be known, as prior information. Generally speaking, prior information is obtained via causal observance, theoretical consideration, or analysis of past data. As established in the work of Kass and Raftery (1995), assumed prior knowledge, as it is applied to current models, are typically selected purely for convenience when there is not enough accurately collected prior data. This can be done because the effect these

assumption have on Bayesian estimations remains small, even when a large sample size is used. The results are helpful when working to use computer modeling with the Gibbs sampler. More specifically, when using the Bayesian approach, it is necessary to evaluate the posterior distribution $[\alpha, \theta, \Omega | Z]$, but the distribution can become relatively complex.

So, in order to correctly demonstrate the characteristics, an increased number of observations are drawn, so that the related empirical distribution of the resulting observations remains consistent with the true distribution. The Gibbs sampler makes an excellent candidate for this process, according to (Geman and Geman, 1984), because it can simulate α, θ and Ω, all from the conditional distribution. However, as a result of the existence of ordered categorical variables in this case, the related conditional distributions can be made too complex to easily derive or simulating data from them. This encourages the additional escalation of Y, the latent matrix, in the posterior analysis, and motivates attention to the joint posterior distribution $[\alpha, \theta, \Omega, Y | Z]$. To garner observations of this posterior distribution, using the Gibbs sampler, it is essential to begin with the starting values $(\alpha^{(0)}, \theta^{(0)}, \Omega^{(0)}, Y^{(0)})$. The following procedure is then implemented to simulate $(\alpha^{(1)}, \theta^{(1)}, \Omega^{(1)}, Y^{(1)})$ and so on. More specifically at the mth reiteration of the current values $\alpha^{(m)}, \theta^{(m)}, \Omega^{(m)}, Y^{(m)}$.

1. Generate $\Omega^{(m+1)}$ from $p(\Omega | \theta^{(m+1)}, \alpha^{(m)}, Y^{(m)}, Z)$

2. Generate $\theta^{(m+1)}$ from $p(\theta | \Omega^{(m+1)}, \alpha^{(m)}, Y^{(m)}, Z)$

3. Generate $(\alpha^{(m+1)}, Y^{(m+1)})$ from $p(\alpha, Y | \theta^{(m+1)}, \Omega^{(m+1)}, Z)$ (9)

The cycle, as previously defined, will give us $(\alpha^{(m+1)}, \theta^{(m+1)}, \Omega^{(m+1)}, Y^{(m+1)})$, only occurring after the m*th* repetition. So, as m approaches infinity, the joint distribution of the value of $(\alpha^{(m)}, \theta^{(m)}, \Omega^{(m)}, Y^{(m)})$ can be proven to move toward the joint posterior distribution $[\alpha, \theta, \Omega, Y | Z]$ (see Geman and Geman, (1984); Geyer, (1992)).

The sequences in which the quantities are replicated from the joint posterior distribution are then used during the calculation of the Bayesian estimates and other similarly related statistics.

4. The Normal Distribution

To indicate that $y^{*} \square N(\mu, \sigma^2)$. y^{*} has the pdf:

$$f(y^{*} | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\Pi}} \exp\left[\frac{-1}{2}\left(\frac{y^{*} - \mu}{\sigma}\right)^2\right]$$

which is plotted in Panel A Figure 1, the cdf is

$$F(y^{*} | \mu, \sigma) = \int_{-\infty}^{y^{*}} f(z | \mu, \sigma) dz = \Pr(Y^{*} \leq y^{*})$$

           (10)

so that

$$\Pr(Y^* > y^*) = 1 - F(y^* \mid \mu, \sigma)$$

$F(c \mid \mu, \sigma)$ is the shaded region in Panel A Figure 1 and $1 - F(c \mid \mu, \sigma)$ is the region to the right of c.

When μ=0 and σ=1, the standard normal distribution is written in the simplified notation:

$$\phi(y^*) = f(y^* \mid \mu = 0, \sigma = 1)$$
$$\Phi(y^*) = F(y^* \mid \mu = 0, \sigma = 1)$$

Any normal distribution, regardless of its mean μ and variance σ^2, can be written as a function of the standard normal distribution. The pdf can be written as

$$f(y^* \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\Pi}} \exp\left[-\frac{1}{2}\left(\frac{y^* - \mu}{\sigma}\right)^2\right] = \frac{1}{\sigma}\phi\left(\frac{y^* - \mu}{\sigma}\right) \tag{11}$$

and the cdf of y^* can be written as

$$\Pr(Y^* \leq y^*) = \Phi\left(\frac{y^* - \mu}{\sigma}\right) \tag{12}$$

so that,

$$\Pr(Y^* > y^*) = 1 - \Phi\left(\frac{y^* - \mu}{\sigma}\right)$$

Hidden continuous normal distribution.

5. The Truncated Normal Distribution

In this section, It is significant to discuss the properties of truncation. Trucation is the effect of data manipulation that occurs any time a sample is drawn from a larger population. In essence the information is altered, or truncated, because only a sample and not a whole population of interest is considered. For example, if you study life-style as it relates to earnings, and you consider income as it relates to the national average, some populations above and below a certain, identified point, may provide insignificant with regard to the whole population or the population about whom inferences are drawn.

As such, truncation can most easily be described as the characteristic a numeric distribution from which a sample is drawn from a restricted segment of the population. Truncated distributions are, thus, simply a segment of an untrucnated distribution which fall within a given range of specified values (see Scott Long, (1997); Wooldridge, (2010)). Statistical applications which base their findings on continuous is the part of an untruncated distribution that is above or below some specified value. More specifically, when values below c are deleted, the variable y|y> c has a truncated normal distribution. In terms of Panel A Figure 1, we want to consider the distribution of y^* in the unshaded

region, while ignoring all cases in the shaded region. The truncated pdf is created by dividing the pdf of the original distribution by the region to the right of c. This forces the resulting distribution to have an area of 1: (Scott Long, 1997)

$$f(y \mid y > c, \mu, \sigma) = \frac{f(y^* \mid \mu, \sigma)}{\Pr(Y^* > c)} \tag{13}$$

The truncated distribution is shown in Panel B Figure 1 by the solid line. The mass of the shaded region has been distributed over the region to the right of c, making the curve slightly higher over this region. This is seen by comparing the solid curve for the truncated distribution to the dotted line for the normal distribution without truncation. Using the results from Equations 12 and 13, we can write the truncated distribution as

$$f(y \mid y > c, \mu, \sigma) = \frac{\frac{1}{\sigma}\phi\left(\frac{y^* - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{c - \mu}{\sigma}\right)} = \frac{\frac{1}{\sigma}\phi\left(\frac{\mu - y^*}{\sigma}\right)}{\Phi\left(\frac{\mu - c}{\sigma}\right)}$$

Given that the left-hand side of the distribution has been truncated, $E(y \mid y > c)$ must be larger than $E(y^*) = \mu$. Specifically, if y^* is normal

$$E(y \mid y > c) = \mu + \sigma \frac{\phi\left(\frac{\mu - c}{\sigma}\right)}{\Phi\left(\frac{\mu - c}{\sigma}\right)} = \mu + \sigma\lambda\left(\frac{\mu - c}{\sigma}\right) \tag{14}$$

where $\lambda(.) = \dfrac{\phi(.)}{\Phi(.)}$

A sampling distribution is truncated if for some reason, we never observe cases above or below a specified point, although in the permissible range of observations the data follow a standard distribution. It is very important to realise that the I(,) construct is not appropriate for truncated distributions with unknown parameters, since the generated likelihood term will ignore the truncation and be incorrect. However, the I(,) construct can be used when specifying truncated prior distributions with no unknown parameters. The indicator I(,) is used for censoring not for truncation because we have unobserved dependent variables in the SEMs and we can't use truncation with this type of variables (Lunn et al., 2012).

6. The Censored Normal Distribution

A data point is a censored observation when we do not know its exact value, but we do know that it lies above or below a point c, say, or within a specified interval.

The relevant distribution theory for censored variables has a lot in common with the distribution theory for truncated variables. It begins with normal distributions, because much of the work is still based on the assumption of normality. It also assumes that the censoring point c exists. Thus, to integrate the distribution to 1, it is scaled according to

the probability that a given observation from the untruncated population falls within the parameters for a population of interest (Greene, 2003).

When a distribution is censored on the left, observations with values at or below c are set to c_y:

$$y = \begin{bmatrix} y^* & if \ y^* > c \\ c_y & if \ y^* \leq c \end{bmatrix}$$

Most often, $c_y = c$, but other values such as zero are also useful. Panel C of Figure 1 plots a censored normal variables, where the censored observation is indicated by the spike at $y = c$. From Equation 12, we know that if y^* is normal, then the probability of an observation being censored is

$$Pr(Censored) = Pr(y^* \leq c) = \Phi\left(\frac{c-\mu}{\sigma}\right)$$

and the probability of a case not being censored is

$$Pr(Uncensored) = Pr(y^* > c) = 1 - \Phi\left(\frac{c-\mu}{\sigma}\right) = \Phi\left(\frac{\mu-c}{\sigma}\right)$$

Thus, the expected value of a censored variable equals

$$E(y) = [Pr(Uncensored) \times E \ (y|y>c)] + [Pr(Censored) \times E \ (y|y=c_y)]$$

$$= \left\{\Phi\left(\frac{\mu-c}{\sigma}\right)\left[\mu + \sigma\lambda\left(\frac{\mu-c}{\sigma}\right)\right]\right\} + \Phi\left(\frac{c-\mu}{\sigma}\right)c_y \qquad (15)$$

where the last equality uses Equation 14 consider how the expected value of the censored value depends on c. As c approaches ∞, the probability of being censored approaches 1 and E(y) approaches the censoring value c_y. As c approaches -∞, the probability of being censored approaches 0 and E(y) approaches the uncensored mean μ (<u>Scott Long, 1997</u>).
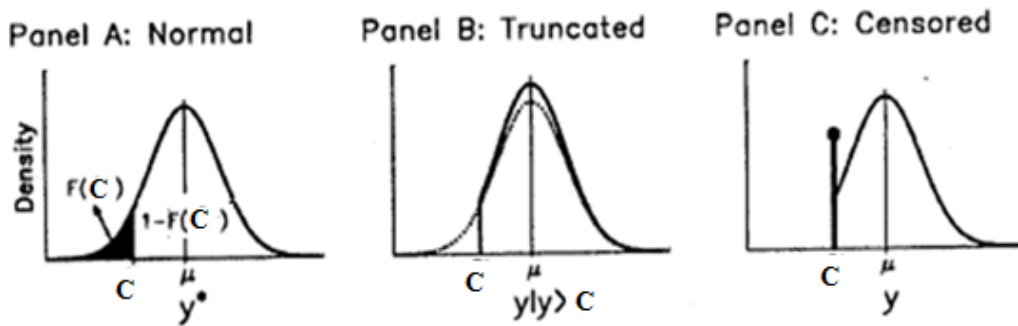

Figure 1. The Normal Distribution with Truncation and Censoring

26

Pak.j.stat.oper.res. Vol.XIII No.1 2017 pp17-45
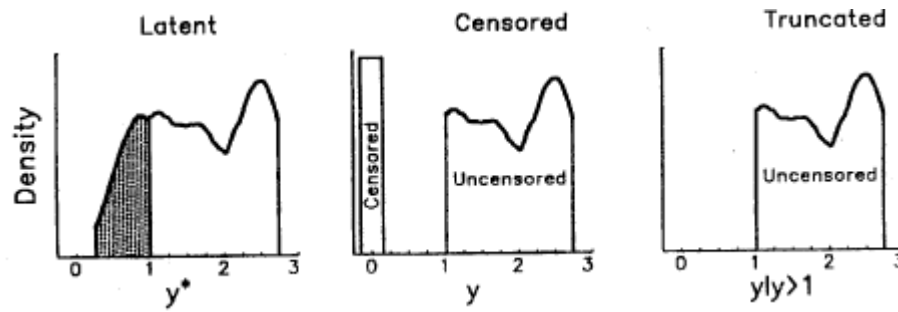
Figure 2. Latent, Censored, and Truncation Variables

7. Model Comparison

A model comparison statistic DIC (see Spiegelhalter *et al.*, 2002) is a generalization of the Akaike Information Criterion (AIC; Akaike, (1973). Under a competing model M_k with a vector of unknown parameters θ_k, the DIC is computed as follows:

$$DIC_k = \overline{D(\theta_k)} + d_k, \tag{16}$$

where $\overline{D(\theta_k)}$ measures the goodness of fit of the model, and is defined as

$$\overline{D(\theta_k)} = \mathrm{E}_{\theta_k}\{-2\log p(Z \mid \theta_k, M_k) \mid Z\}. \tag{17}$$

Here, d_k is the effective number of parameters in M_k, and is defined as

$$d_k = \mathrm{E}_{\theta_k}[-2\log p(Z \mid \theta_k, M_k) \mid Z] + 2\log f(Z \mid \tilde{\theta}). \tag{18}$$

in which $\tilde{\theta}$ is the Bayesian estimate of θ. Let $\{\theta_k^{(t)} : t = 1,...,T\}$ be a sample of observations simulated from the posterior distribution. The expectations in Equations (17) and (18) can be estimated as follows:

$$\mathrm{E}_{\theta_k}\{-2\log p(Z \mid \theta_k, M_k) \mid Z\} = -\frac{2}{T}\sum_{t=1}^{T}\log p(Z \mid \theta_k^{(t)}, M_k). \tag{19}$$

In Bayesian SEMs, the model with the smaller DIC value is selected.

8. Real Example

The data used for modeling were two independent samples selected from the natural history studies on rural drug use practices in Ohio (n=200) and Kentucky (n=200) in the USA between 2003 and 2005 (Booth et al., 2006). Study eligibility included the following: (1) having provided informed consent to participate in the study; (2) being aged 18 or older; (3) self-reported use of crack-cocaine, powdered cocaine, or methamphetamine/amphetamine in the past 30 days; (4) no recent formal substance abuse treatment (past 30 days); and (5) residence in one of the targeted counties. Respondent-driven sampling, which has been increasingly applied to recruit hidden drug user populations, was used for sample recruitment in the current project. The project

participants were followed up every six months after the baseline interview for three consecutive years (Wang & Wang, 2012).

Again, the BSI-18 scale was used for model demonstration. The three dimensions of psychiatric disorders measured by the BSI-18 are: somatization (SOM), depression (DEP), and anxiety (ANX). Each of the three subscales was measured by six items, respectively. All the BSI-18 items are measured on a fivepoint Likert scale (1, not at all; 2, a little bit; 3, moderately; 4, quite a bit; 5, extremely). All subscales of the BSI-18 have excellent reliabilities with Cronbach's alpha greater than 0.80 in the populations under study.

9. Analysis of Real Data

A real data study is presented here to give some idea of the empirical performance of the proposed Bayesian approach in which 18 manifest variables are related to two basic latent variables $(\eta, \xi_{i1}, \xi_{i2})$ from multiple group nonlinear SEMs defined in Equation 21 and Equation 22 respectively. Hence, some quadratic and interaction effects of the latent variables are considered. To illustrate the Bayesian methods in analysing linear and nonlinear structural equation models with ordered categorical variables, we use a simulated data set that is related to random vectors with G=1,2, $z_i^{(g)} = (z_{i1}^{(g)}, z_{i2}^{(g)}, ..., z_{i18}^{(g)})'$, let $y_i^{(g)} = (y_{i1}^{(g)}, y_{i2}^{(g)}, ..., y_{i18}^{(g)})'$ be the latent continuous random vector which corresponds to the ordinal variables $z_{i1}^{(g)}, z_{i2}^{(g)}, ..., z_{i18}^{(g)}$ where $z_i^{(g)}, i=1,...,n$ are ordered categorical variables that are related to (3) latent variables $w_i^{(g)} = (\eta_i^{(g)}, \xi_{i1}^{(g)}, \xi_{i2}^{(g)})$, $\varepsilon_i^{(g)} = (\varepsilon_{i1}^{(g)}, \varepsilon_{i2}^{(g)}, ..., \varepsilon_{i18}^{(g)})$, with the following values of the parameters in $\mu^{(g)} = (\mu_1^{(g)}, \mu_2^{(g)}, ..., \mu_{18}^{(g)})'$ and $\Lambda^{(g)} = (\lambda_1^{(g)'}, \lambda_2^{(g)'}, ..., \lambda_{15}^{(g)'})'$

$$
\Lambda^{(g)} = \Lambda^{(1)} \begin{bmatrix} 1^* & \lambda_{21}^{(1)} & \lambda_{31}^{(1)} & \lambda_{41}^{(1)} & \lambda_{51}^{(1)} & \lambda_{61}^{(1)} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1 & \lambda_{82}^{(1)} & \lambda_{92}^{(1)} & \lambda_{102}^{(1)} & \lambda_{112}^{(1)} & \lambda_{122}^{(1)} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1 & \lambda_{143}^{(1)} & \lambda_{153}^{(1)} & \lambda_{163}^{(1)} & \lambda_{173}^{(1)} & \lambda_{183}^{(1)} \end{bmatrix},
$$

$$
\Lambda^{(2)} \begin{bmatrix} 1^* & \lambda_{21}^{(2)} & \lambda_{31}^{(2)} & \lambda_{41}^{(2)} & \lambda_{51}^{(2)} & \lambda_{61}^{(2)} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1 & \lambda_{82}^{(2)} & \lambda_{92}^{(2)} & \lambda_{102}^{(2)} & \lambda_{112}^{(2)} & \lambda_{122}^{(2)} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1 & \lambda_{143}^{(2)} & \lambda_{153}^{(2)} & \lambda_{163}^{(2)} & \lambda_{173}^{(2)} & \lambda_{183}^{(2)} \end{bmatrix}.
$$

$$
\Phi^{(g)} = \Phi^{(1)} \begin{bmatrix} \phi_{11}^{(1)} & \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{bmatrix}, \quad \Phi^{(2)} \begin{bmatrix} \phi_{11}^{(2)} & \\ \phi_{21}^{(2)} & \phi_{22}^{(2)} \end{bmatrix}. \tag{20}
$$

Where parameters with an asterisk are treated as fixed for identifying the model.

The relationships of the latent variables in $w_i = (\eta_i, \xi_{1i}, \xi_{2i})$ are assessed by the nonlinear structural equation which described in equation.

$$
\eta_i^{(1)} = \gamma_1^{(1)} \xi_{i1}^{(1)} + \gamma_2^{(1)} \xi_{i2}^{(1)} + \gamma_3^{(1)} \xi_{i1}^{(1)} \xi_{i2}^{(1)} + \gamma_4^{(1)} \xi_{i1}^{(1)} \xi_{i1}^{(1)} + \gamma_5^{(1)} \xi_{i2}^{(1)} \xi_{i2}^{(1)} + \delta_i^{(1)}, \tag{21}
$$

$$
\eta_i^{(2)} = \gamma_1^{(2)} \xi_{i1}^{(2)} + \gamma_2^{(2)} \xi_{i2}^{(2)} + \gamma_3^{(2)} \xi_{i1}^{(2)} \xi_{i2}^{(2)} + \gamma_4^{(2)} \xi_{i1}^{(2)} \xi_{i1}^{(2)} + \gamma_5^{(2)} \xi_{i2}^{(2)} \xi_{i2}^{(2)} + \delta_i^{(2)}. \tag{22}
$$

The following accurate prior inputs of the hyperparameter values in the conjugate prior distributions of the parameters are considered:

Prior I. Elements in μ_0, λ_{0k} and $\lambda_{0\xi k}$ in Equations (8) are set equal to 1;

$R_0^{-1} = 8\Phi$ H_{0u}, H_{0k} and $H_{0\xi k}$ are taken to be 0.25 times the identity matrices; $\alpha_{0k} = 10$, $\beta_{0k} = 8$, $\rho_0 = 30$.

Prior II. Elements in μ_0, λ_{0k} and $\lambda_{0\xi k}$ in Equations (8) are set equal to 0.5;

$R_0^{-1} = 8\Phi$ H_{0u}, H_{0k} and $H_{0\xi k}$ are taken to be 0.25 times the identity matrices; $\alpha_{0k} = 10$, $\beta_{0k} = 8, \rho_0 = 30$.

The prior is informative and can have a significant effect on the parameter estimates for a small sample size case.

A data set (n_1=200, n_2=200) was analyzed by WinBUGS. In comparing the Bayesian analyses of structural equation models with data, the MCMC procedure for analysing data required more iterations to converge. Bayesian estimates were obtained from T=10000. Iterations after discarding 4000 burn-in iterations with ordered categorical variables. The WinBUGS software (Spiegelhalter *et al.*, 2003) can produce Bayesianestimates of the parameters in some two-level nonlinear SEMs. To demonstrate this, we apply WinBUGS to analyse the current aids data based on (Model 5) with different prior inputs.

10. Results and Discussion

The objective of this section is to present results of a simulation study to reveal the empirical performances of the Bayesian estimates and the DIC for model comparison. For linear and nonlinear SEMs we have the following proposed models:

$$Model\ 1: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \gamma_4\xi_{i1}\xi_{i2} + \delta_i,$$
$$Model\ 2: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \gamma_4\xi_{i1}\xi_{i3} + \delta_i,$$
$$Model\ 3: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \gamma_4\xi_{i2}\xi_{i3} + \delta_i, \tag{23}$$
$$Model\ 4: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \gamma_4\xi_{i1}\xi_{i2} + \gamma_5\xi_{i2}^2 + \delta_i,$$
$$Model\ 5: \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \gamma_4\xi_{i1}\xi_{i2} + \gamma_5\gamma_4\xi_{i1}\xi_{i1} + \gamma_4\xi_{i2}\xi_{i3} + \delta_i,$$

The Bayesian estimates of the unknown parameters and the Bayesian model selection statistic DIC are obtained using recently developed powerful tools in statistical computing. All the computational work can be accomplished via the recently developed and freely available software WinBUGS. The purpose of this analysis is to use Bayesian nonlinear multiple group SEMs with ordered categorical data. There are some limitations of the current analysis. First, due to the design of questionnaires and the nature of the problems in behavioural, educational, medical and social sciences, data are often coming from ordered categorical variables with observations in discrete form. In analysing ordered categorical data, the basic assumption in SEM that the data come from a continuous normal distribution is clearly violated, and rigorous analysis that takes into

account the ordered categorical nature is necessary. Hence, clearly, routinely treating ordered categorical variables as normal may lead to erroneous conclusions (see Lee et al., 1990; Olsson, 1979).

A better approach for assessing this kind of discrete data is to treat them as observations obtained from a hidden continuous normal distribution (censored normal distribution) with a threshold specification. Second, the current analysis was conducted under the normality assumption of the observed variables in the model. However, this assumption is likely to be violated. Developing a linear & nonlinear Bayesian approach with hidden continuous normal distribution (right censored normal distribution, left censored normal distribution) to relax the normality assumption in SEMs may represent a future research topic.

Figure 4. Two chains of observation corresponding to (a) $\mu_3^{(2)}$; (b) $\lambda_5^{(1)}$; (c) $\gamma_1^{(2)}$; (d) $\Phi_{22}^{(1)}$ for Multiple SEMs with ordered categorical variables.
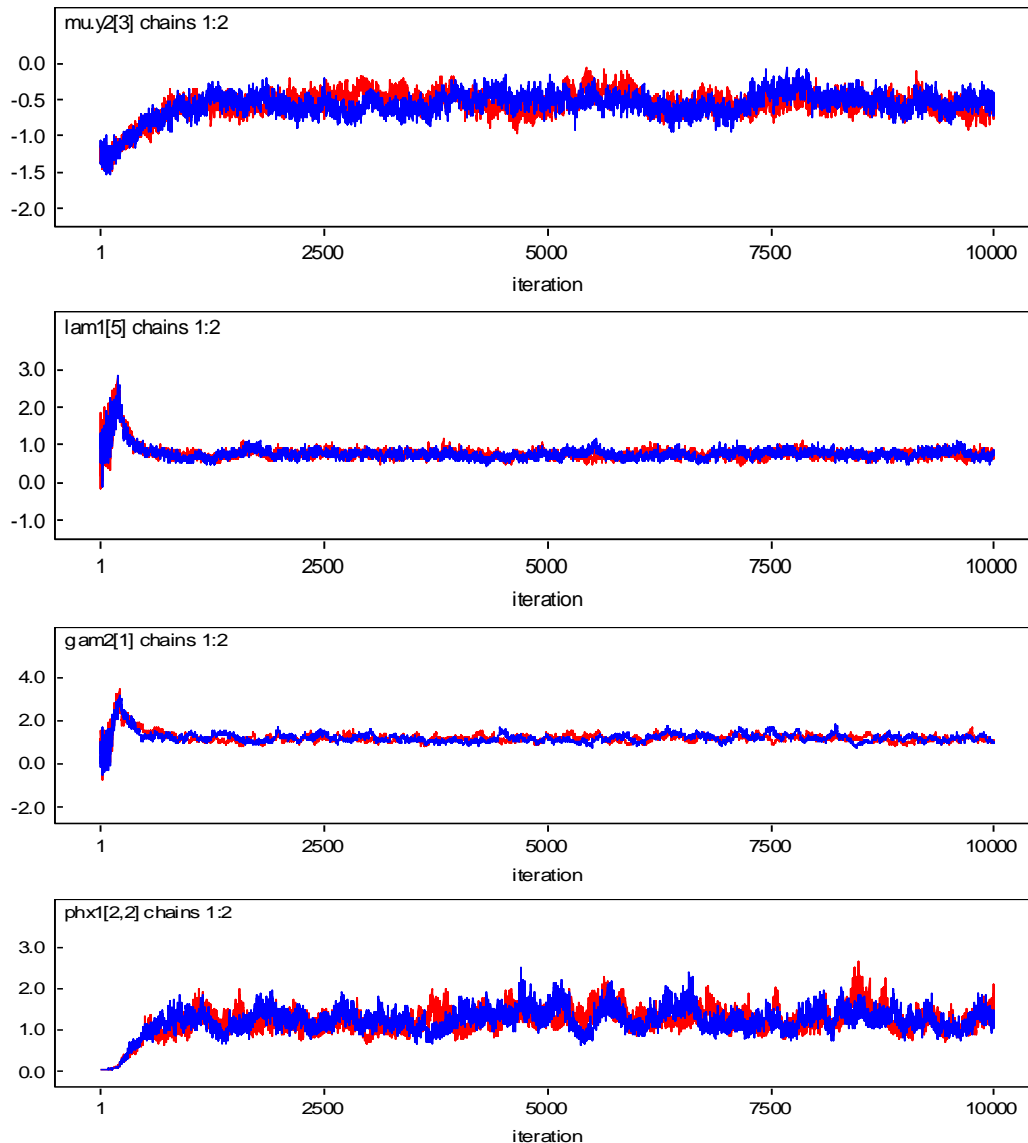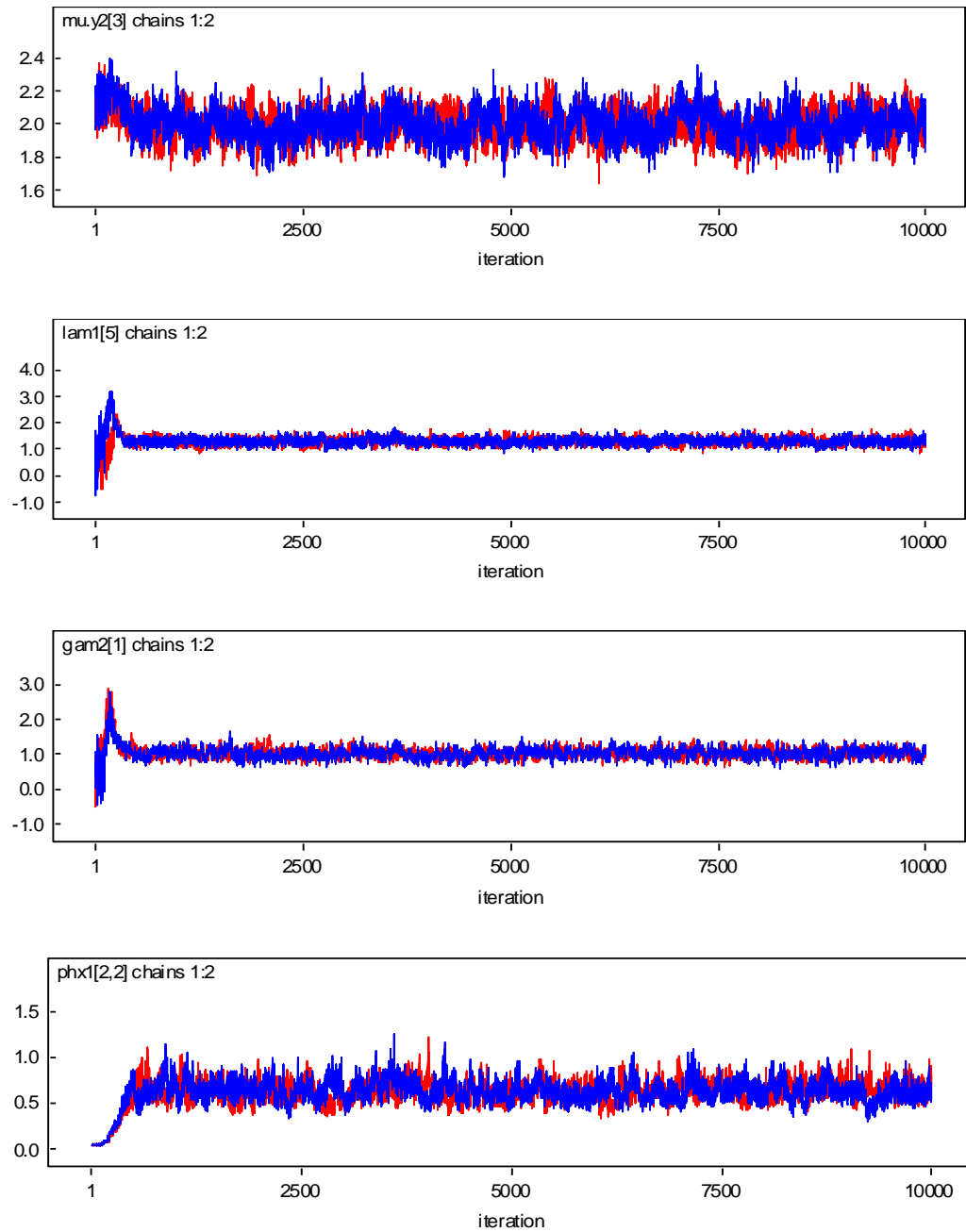
Figure 5. Two chains of observation corresponding to (a) $\mu_3^{(2)}$; (b) $\lambda_5^{(1)}$; (c) $\gamma_1^{(2)}$; (d) $\Phi_{22}^{(1)}$ for Multiple SEMs with ordered categorical variables treated as a continuous normal.

Table 1: **Bayesian Estimation of Multiple Group Nonlinear SEM with Ordered Categorical Variables of First Group under Prior I**

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\mu_1^{(1)}$	-2.221	0.13	[-2.481, -1.967]	$\lambda_{51}^{(1)}$	0.905	0.105	[0.708, 1.127]
$\mu_2^{(1)}$	-0.319	0.108	[-0.527, -0.097]	$\lambda_{61}^{(1)}$	0.724	0.087	[0.566, 0.901]
$\mu_3^{(1)}$	-0.612	0.089	[-0.779, -0.433]	$\lambda_{82}^{(1)}$	0.981	0.115	[0.777, 1.229]
$\mu_4^{(1)}$	-1.475	0.112	[-1.702, -1.255]	$\lambda_{92}^{(1)}$	1.113	0.156	[0.844, 1.448]
$\mu_5^{(1)}$	-0.499	0.106	[-0.704, -0.283]	$\lambda_{102}^{(1)}$	1.028	0.149	[0.762, 1.341]
$\mu_6^{(1)}$	-0.728	0.084	[-0.888, -0.556]	$\lambda_{112}^{(1)}$	1.180	0.141	[0.921, 1.478]
$\mu_7^{(1)}$	-0.697	0.115	[-0.914, -0.465]	$\lambda_{122}^{(1)}$	1.735	0.227	[1.322, 2.209]
$\mu_8^{(1)}$	-0.779	0.105	[-0.973, -0.564]	$\lambda_{143}^{(1)}$	0.684	0.105	[0.492, 0.906]
$\mu_9^{(1)}$	-1.572	0.138	[-1.845, -1.302]	$\lambda_{153}^{(1)}$	0.783	0.107	[0.591, 1.012]
$\mu_{10}^{(1)}$	-1.859	0.131	[-2.116, -1.597]	$\lambda_{163}^{(1)}$	0.836	0.110	[0.631, 1.057]
$\mu_{11}^{(1)}$	-1.196	0.126	[-1.436, -0.943]	$\lambda_{173}^{(1)}$	0.483	0.133	[0.255, 0.758]
$\mu_{12}^{(1)}$	-2.067	0.182	[-2.419, -1.708]	$\lambda_{183}^{(1)}$	0.796	0.127	[0.569, 1.081]
$\mu_{13}^{(1)}$	-1.181	0.161	[-1.489, -0.858]	$\phi_{11}^{(1)}$	1.473	0.266	[0.999, 2.044]
$\mu_{14}^{(1)}$	-0.896	0.139	[-1.158, -0.618]	$\phi_{12}^{(1)}$	1.285	0.218	[0.903, 1.756]
$\mu_{15}^{(1)}$	-0.53	0.141	[-0.799, -0.253]	$\phi_{22}^{(1)}$	1.361	0.274	[0.906, 1.972]
$\mu_{16}^{(1)}$	-1.115	0.132	[-1.359, -0.845]	$\gamma_1^{(1)}$	0.411	0.205	[0.033, 0.816]
$\mu_{17}^{(1)}$	-2.305	0.157	[-2.635, -2.022]	$\gamma_2^{(1)}$	0.788	0.215	[0.339, 1.206]
$\mu_{18}^{(1)}$	-1.812	0.135	[-2.075, -1.548]	$\gamma_3^{(1)}$	0.278	0.407	[-0.551, 1.037]
$\lambda_{21}^{(1)}$	0.979	0.111	[0.776, 1.218]	$\gamma_4^{(1)}$	0.295-	0.208	[-0.724, 0.131]
$\lambda_{31}^{(1)}$	0.709	0.088	[0.546, 0.893]	$\gamma_5^{(1)}$	0.186-	0.263	[-0.725, 0.391]
$\lambda_{41}^{(1)}$	0.761	0.119	[0.548, 1.016]	$\psi_{\varepsilon\delta}^{(1)}$	0.381	0.072	[0.264, 0.551]

*HPD: highest posterior density

Table 2: **Bayesian Estimation of Multiple Group Nonlinear SEM with Ordered Categorical Variables of Second Group under Prior I**

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\mu_1^{(2)}$	-2.379	0.139	[-2.642, -2.105]	$\lambda_{51}^{(2)}$	0.898	0.114	[0.697,1.140]
$\mu_2^{(2)}$	-0.158	0.112	[-0.373, 0.066]	$\lambda_{61}^{(2)}$	0.794	0.094	[0.628,0.996]
$\mu_3^{(2)}$	-0.552	0.108	[-0.754, -0.331]	$\lambda_{82}^{(2)}$	1.088	0.117	[0.871,1.326]
$\mu_4^{(2)}$	-1.520	0.120	[-1.754, -1.280]	$\lambda_{92}^{(2)}$	1.388	0.177	[1.071,1.762]
$\mu_5^{(2)}$	-0.473	0.125	[-0.711, -0.218]	$\lambda_{102}^{(2)}$	0.826	0.130	[0.595,1.102]
$\mu_6^{(2)}$	-0.742	0.097	[-0.925, -0.544]	$\lambda_{112}^{(2)}$	1.464	0.164	[1.163,1.822]
$\mu_7^{(2)}$	-0.853	0.118	[-1.078, -0.615]	$\lambda_{122}^{(2)}$	1.564	0.187	[1.235,1.964]
$\mu_8^{(2)}$	-0.494	0.128	[-0.736, -0.240]	$\lambda_{143}^{(2)}$	1.098	0.139	[0.846,1.391]
$\mu_9^{(2)}$	-1.429	0.157	[-1.752, -1.131]	$\lambda_{153}^{(2)}$	1.017	0.115	[0.800,1.252]
$\mu_{10}^{(2)}$	-2.047	0.132	[-2.307, -1.792]	$\lambda_{163}^{(2)}$	0.841	0.089	[0.678,1.032]
$\mu_{11}^{(2)}$	-1.245	0.149	[-1.526, -0.940]	$\lambda_{173}^{(2)}$	0.852	0.164	[0.553,1.197]
$\mu_{12}^{(2)}$	-2.229	0.169	[-2.561, -1.894]	$\lambda_{183}^{(2)}$	0.859	0.127	[0.634,1.141]
$\mu_{13}^{(2)}$	-1.174	0.172	[-1.506, -0.832]	$\phi_{11}^{(2)}$	2.014	0.403	[1.319,2.879]
$\mu_{14}^{(2)}$	-0.726	0.197	[-1.109, -0.331]	$\phi_{12}^{(2)}$	1.752	0.275	[1.280,2.368]
$\mu_{15}^{(2)}$	-0.710	0.168	[-1.034, -0.377]	$\phi_{22}^{(2)}$	1.753	0.287	[1.238,2.352]
$\mu_{16}^{(2)}$	-1.006	0.137	[-1.263, -0.725]	$\gamma_1^{(2)}$	1.178	0.148	[0.929,1.511]
$\mu_{17}^{(2)}$	-2.409	0.206	[-2.825, -2.009]	$\gamma_2^{(2)}$	0.142	0.078	[-0.010,0.296]
$\mu_{18}^{(2)}$	-1.817	0.149	[-2.105, -1.512]	$\gamma_3^{(2)}$	0.093	0.325	[-0.663,0.743]
$\lambda_{21}^{(2)}$	0.981	0.112	[0.784,1.219]	$\gamma_4^{(2)}$	0.005	0.240	[-0.419,0.529]
$\lambda_{31}^{(2)}$	0.922	0.107	[0.735,1.154]	$\gamma_5^{(2)}$	-0.171	0.186	[-0.577,0.185]
$\lambda_{41}^{(2)}$	0.740	0.120	[0.529,0.994]	$\psi_{\varepsilon\delta}^{(2)}$	0.386	0.070	[0.271,0.546]

Table 3: **Bayesian Estimation of Multiple Group Nonlinear SEM with Ordered Categorical Variables of First Group under Prior II**

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\mu_1^{(1)}$	-2.234	0.128	[-2.487, -1.981]	$\lambda_{51}^{(1)}$	0.924	0.117	[0.718,1.178]
$\mu_2^{(1)}$	-0.334	0.107	[-0.532, -0.113]	$\lambda_{61}^{(1)}$	0.741	0.096	[0.567,0.943]
$\mu_3^{(1)}$	-0.620	0.089	[-0.790, -0.437]	$\lambda_{82}^{(1)}$	1.008	0.111	[0.812,1.242]
$\mu_4^{(1)}$	-1.486	0.117	[-1.717, -1.258]	$\lambda_{92}^{(1)}$	1.140	0.155	[0.856,1.464]
$\mu_5^{(1)}$	-0.513	0.107	[-0.716, -0.293]	$\lambda_{102}^{(1)}$	1.048	0.150	[0.775,1.364]
$\mu_6^{(1)}$	-0.737	0.082	[-0.887, -0.565]	$\lambda_{112}^{(1)}$	1.208	0.147	[0.933,1.520]
$\mu_7^{(1)}$	-0.720	0.111	[-0.931, -0.500]	$\lambda_{122}^{(1)}$	1.753	0.209	[1.364,2.206]
$\mu_8^{(1)}$	-0.794	0.105	[-0.992, -0.576]	$\lambda_{143}^{(1)}$	0.709	0.110	[0.505,0.939]
$\mu_9^{(1)}$	-1.585	0.140	[-1.863, -1.308]	$\lambda_{153}^{(1)}$	0.806	0.113	[0.595,1.040]
$\mu_{10}^{(1)}$	-1.873	0.130	[-2.134, -1.614]	$\lambda_{163}^{(1)}$	0.861	0.109	[0.660,1.091]
$\mu_{11}^{(1)}$	-1.211	0.125	[-1.446, -0.947]	$\lambda_{173}^{(1)}$	0.474	0.134	[0.245,0.768]
$\mu_{12}^{(1)}$	-2.097	0.186	[-2.463, -1.717]	$\lambda_{183}^{(1)}$	0.809	0.129	[0.575,1.074]
$\mu_{13}^{(1)}$	-1.216	0.155	[-1.516, -0.907]	$\phi_{11}^{(1)}$	1.404	0.279	[0.906,1.983]
$\mu_{14}^{(1)}$	-0.910	0.137	[-1.173, -0.638]	$\phi_{12}^{(1)}$	1.204	0.216	[0.829,1.663]
$\mu_{15}^{(1)}$	-0.550	0.139	[-0.813, -0.273]	$\phi_{22}^{(1)}$	1.263	0.232	[0.860,1.780]
$\mu_{16}^{(1)}$	-1.139	0.130	[-1.385, -0.888]	$\gamma_1^{(1)}$	0.420	0.192	[0.056, 0.812]
$\mu_{17}^{(1)}$	-2.310	0.155	[-2.641, -2.024]	$\gamma_2^{(1)}$	0.782	0.216	[0.357, 1.206]
$\mu_{18}^{(1)}$	-1.838	0.133	[-2.096, -1.577]	$\gamma_3^{(1)}$	0.140	0.303	[-0.542, 0.753]
$\lambda_{21}^{(1)}$	0.997	0.119	[0.784, 1.253]	$\gamma_4^{(1)}$	-0.235	0.185	[-0.606, 0.130]
$\lambda_{31}^{(1)}$	0.724	0.094	[0.556, 0.922]	$\gamma_5^{(1)}$	-0.113	0.190	[-0.509, 0.275]
$\lambda_{41}^{(1)}$	0.787	0.129	[0.561, 1.061]	$\psi_{\varepsilon\delta}^{(1)}$	0.378	0.071	[0.261,0.539]

Table 4: **Bayesian Estimation of Multiple Group Nonlinear SEM with Ordered Categorical Variables of Second Group under Prior II**

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\mu_1^{(1)}$	-2.370	0.146	[-2.673, -2.097]	$\lambda_{51}^{(1)}$	0.921	0.111	[0.716, 1.153]
$\mu_2^{(1)}$	-0.139	0.126	[-0.384, 0.107]	$\lambda_{61}^{(1)}$	0.814	0.092	[0.652, 1.010]
$\mu_3^{(1)}$	-0.536	0.114	[-0.766, -0.320]	$\lambda_{82}^{(1)}$	1.089	0.115	[0.872, 1.323]
$\mu_4^{(1)}$	-1.511	0.119	[-1.755, -1.286]	$\lambda_{92}^{(1)}$	1.362	0.166	[1.062, 1.718]
$\mu_5^{(1)}$	-0.452	0.133	[-0.708, -0.187]	$\lambda_{102}^{(1)}$	0.826	0.131	[0.588, 1.106]
$\mu_6^{(1)}$	-0.723	0.108	[-0.939, -0.511]	$\lambda_{112}^{(1)}$	1.460	0.152	[1.180, 1.767]
$\mu_7^{(1)}$	-0.829	0.135	[-1.102, -0.579]	$\lambda_{122}^{(1)}$	1.581	0.185	[1.254, 1.975]
$\mu_8^{(1)}$	-0.474	0.140	[-0.752, -0.205]	$\lambda_{143}^{(1)}$	1.100	0.133	[0.860, 1.378]
$\mu_9^{(1)}$	-1.400	0.173	[-1.751, -1.073]	$\lambda_{153}^{(1)}$	1.035	0.116	[0.834, 1.291]
$\mu_{10}^{(1)}$	-2.038	0.143	[-2.335, -1.768]	$\lambda_{163}^{(1)}$	0.845	0.094	[0.673, 1.049]
$\mu_{11}^{(1)}$	-1.219	0.172	[-1.563, -0.892]	$\lambda_{173}^{(1)}$	0.851	0.167	[0.553, 1.211]
$\mu_{12}^{(1)}$	-2.203	0.186	[-2.574, -1.862]	$\lambda_{183}^{(1)}$	0.856	0.118	[0.648, 1.111]
$\mu_{13}^{(1)}$	-1.167	0.168	[-1.490, -0.832]	$\phi_{11}^{(1)}$	1.972	0.415	[1.321, 2.919]
$\mu_{14}^{(1)}$	-0.713	0.203	[-1.102, -0.292]	$\phi_{12}^{(1)}$	1.757	0.322	[1.217, 2.498]
$\mu_{15}^{(1)}$	-0.686	0.178	[-1.020, -0.322]	$\phi_{22}^{(1)}$	1.794	0.347	[1.213, 2.551]
$\mu_{16}^{(1)}$	-0.995	0.141	[-1.267, -0.712]	$\gamma_1^{(1)}$	1.194	0.145	[0.915, 1.484]
$\mu_{17}^{(1)}$	-2.389	0.204	[-2.812, -2.023]	$\gamma_2^{(1)}$	0.139	0.078	[-0.012, 0.299]
$\mu_{18}^{(1)}$	-1.799	0.154	[-2.102, -1.499]	$\gamma_3^{(1)}$	0.096	0.346	[-0.566, 0.642]
$\lambda_{21}^{(1)}$	1.006	0.106	[0.802, 1.218]	$\gamma_4^{(1)}$	0.004	0.230	[-0.417, 0.465]
$\lambda_{31}^{(1)}$	0.942	0.103	[0.752, 1.157]	$\gamma_5^{(1)}$	-0.166	0.242	[-0.584, 0.297]
$\lambda_{41}^{(1)}$	0.753	0.111	[0.553, 0.981]	$\psi_{\varepsilon\delta}^{(1)}$	0.383	0.071	[0.266, 0.541]

Table 5: **Bayesian Estimation of Multiple Group Nonlinear SEM with Ordered Categorical Variables Treated as a Continuous Normal of First Group under Prior I**

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\mu_1^{(1)}$	1.468	0.073	[1.326, 1.614]	$\lambda_{51}^{(1)}$	1.326	0.128	[1.082,1.581]
$\mu_2^{(1)}$	2.068	0.090	[1.890, 2.246]	$\lambda_{61}^{(1)}$	1.288	0.127	[1.051,1.544]
$\mu_3^{(1)}$	2.232	0.085	[2.064, 2.396]	$\lambda_{82}^{(1)}$	1.092	0.107	[0.896,1.316]
$\mu_4^{(1)}$	1.624	0.079	[1.469, 1.780]	$\lambda_{92}^{(1)}$	0.765	0.095	[0.589,0.964]
$\mu_5^{(1)}$	2.183	0.094	[1.996, 2.366]	$\lambda_{102}^{(1)}$	0.744	0.094	[0.563,0.934]
$\mu_6^{(1)}$	2.319	0.091	[2.137, 2.498]	$\lambda_{112}^{(1)}$	1.059	0.111	[0.851,1.281]
$\mu_7^{(1)}$	1.929	0.088	[1.754, 2.103]	$\lambda_{122}^{(1)}$	0.898	0.102	[0.707,1.110]
$\mu_8^{(1)}$	2.053	0.083	[1.893, 2.220]	$\lambda_{143}^{(1)}$	0.715	0.099	[0.528,0.918]
$\mu_9^{(1)}$	1.457	0.072	[1.314, 1.600]	$\lambda_{153}^{(1)}$	0.891	0.100	[0.700,1.095]
$\mu_{10}^{(1)}$	1.514	0.072	[1.375, 1.659]	$\lambda_{163}^{(1)}$	0.986	0.096	[0.806,1.181]
$\mu_{11}^{(1)}$	1.820	0.084	[1.656, 1.988]	$\lambda_{173}^{(1)}$	0.180	0.049	[0.086,0.279]
$\mu_{12}^{(1)}$	1.471	0.078	[1.318, 1.622]	$\lambda_{183}^{(1)}$	0.702	0.084	[0.543,0.875]
$\mu_{13}^{(1)}$	1.815	0.103	[1.610, 2.024]	$\phi_{11}^{(1)}$	0.473	0.070	[0.345,0.624]
$\mu_{14}^{(1)}$	2.011	0.094	[1.826, 2.197]	$\phi_{12}^{(1)}$	0.446	0.067	[0.327,0.588]
$\mu_{15}^{(1)}$	2.279	0.099	[2.080, 2.472]	$\phi_{22}^{(1)}$	0.620	0.097	[0.445,0.827]
$\mu_{16}^{(1)}$	1.942	0.099	[1.746, 2.142]	$\gamma_1^{(1)}$	0.572	0.186	[0.203, 0.933]
$\mu_{17}^{(1)}$	1.122	0.044	[1.036, 1.209]	$\gamma_2^{(1)}$	0.537	0.179	[0.205, 0.913]
$\mu_{18}^{(1)}$	1.517	0.082	[1.353, 1.677]	$\gamma_3^{(1)}$	-0.177	0.305	[-0.782, 0.392]
$\lambda_{21}^{(1)}$	1.365	0.121	[1.136, 1.606]	$\gamma_4^{(1)}$	0.033	0.225	[-0.398, 0.470]
$\lambda_{31}^{(1)}$	1.155	0.117	[0.934, 1.394]	$\gamma_5^{(1)}$	0.133	0.171	[-0.180, 0.484]
$\lambda_{41}^{(1)}$	0.808	0.113	[0.594, 1.036]	$\psi_{\varepsilon\delta}^{(1)}$	0.263	0.038	[0.195,0.349]

Table 6: **Bayesian Estimation of Multiple Group Nonlinear SEM with Ordered Categorical Variables Treated as a Continuous Normal of Second Group under Prior I**

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\mu_1^{(2)}$	1.321	0.061	[1.203,1.442]	$\lambda_{51}^{(2)}$	1.372	0.139	[1.104, 1.653]
$\mu_2^{(2)}$	2.026	0.080	[1.870,2.181]	$\lambda_{61}^{(2)}$	1.480	0.131	[1.234, 1.748]
$\mu_3^{(2)}$	1.999	0.082	[1.838,2.157]	$\lambda_{82}^{(2)}$	1.168	0.118	[0.945, 1.415]
$\mu_4^{(2)}$	1.485	0.068	[1.355,1.617]	$\lambda_{92}^{(2)}$	0.937	0.094	[0.761, 1.127]
$\mu_5^{(2)}$	2.083	0.087	[1.911,2.253]	$\lambda_{102}^{(2)}$	0.594	0.081	[0.441, 0.757]
$\mu_6^{(2)}$	2.062	0.083	[1.902,2.222]	$\lambda_{112}^{(2)}$	1.182	0.107	[0.981, 1.400]
$\mu_7^{(2)}$	1.719	0.077	[1.572,1.871]	$\lambda_{122}^{(2)}$	0.798	0.083	[0.643, 0.974]
$\mu_8^{(2)}$	2.125	0.084	[1.958,2.291]	$\lambda_{143}^{(2)}$	0.876	0.089	[0.710, 1.054]
$\mu_9^{(2)}$	1.411	0.070	[1.271,1.548]	$\lambda_{153}^{(2)}$	0.931	0.084	[0.772, 1.099]
$\mu_{10}^{(2)}$	1.378	0.060	[1.258,1.495]	$\lambda_{163}^{(2)}$	0.972	0.078	[0.827, 1.131]
$\mu_{11}^{(2)}$	1.624	0.078	[1.473,1.776]	$\lambda_{173}^{(2)}$	0.276	0.052	[0.176, 0.380]
$\mu_{12}^{(2)}$	1.308	0.060	[1.189,1.428]	$\lambda_{183}^{(2)}$	0.683	0.069	[0.551, 0.823]
$\mu_{13}^{(2)}$	1.554	0.085	[1.387,1.721]	$\phi_{11}^{(2)}$	0.412	0.061	[0.307, 0.546]
$\mu_{14}^{(2)}$	1.721	0.092	[1.542,1.902]	$\phi_{12}^{(2)}$	0.416	0.060	[0.311, 0.544]
$\mu_{15}^{(2)}$	1.787	0.086	[1.618,1.954]	$\phi_{22}^{(2)}$	0.617	0.096	[0.446, 0.827]
$\mu_{16}^{(2)}$	1.752	0.083	[1.589,1.913]	$\gamma_1^{(2)}$	1.021	0.129	[0.782, 1.296]
$\mu_{17}^{(2)}$	1.069	0.050	[0.971,1.167]	$\gamma_2^{(2)}$	0.013	0.062	[-0.110, 0.135]
$\mu_{18}^{(2)}$	1.332	0.071	[1.194,1.473]	$\gamma_3^{(2)}$	-0.192	0.331	[-0.767, 0.444]
$\lambda_{21}^{(2)}$	1.517	0.128	[1.279,1.781]	$\gamma_4^{(2)}$	0.242	0.262	[-0.283, 0.742]
$\lambda_{31}^{(2)}$	1.524	0.129	[1.278,1.784]	$\gamma_5^{(2)}$	0.257	0.184	[-0.100, 0.578]
$\lambda_{41}^{(2)}$	0.922	0.109	[0.716,1.141]	$\psi_{\varepsilon\delta}^{(2)}$	0.236	0.032	[0.180, 0.305]

Table 7: **Bayesian Estimation of Multiple Group Nonlinear SEM with Ordered Categorical Variables Treated as a Continuous Normal of First Group under Prior II**

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\mu_1^{(1)}$	1.470	0.072	[1.329, 1.611]	$\lambda_{51}^{(1)}$	1.332	0.129	[1.090, 1.595]
$\mu_2^{(1)}$	2.070	0.088	[1.897, 2.247]	$\lambda_{61}^{(1)}$	1.293	0.127	[1.058, 1.557]
$\mu_3^{(1)}$	2.234	0.085	[2.066, 2.400]	$\lambda_{82}^{(1)}$	1.097	0.107	[0.898, 1.311]
$\mu_4^{(1)}$	1.626	0.079	[1.470, 1.780]	$\lambda_{92}^{(1)}$	0.768	0.094	[0.591, 0.959]
$\mu_5^{(1)}$	2.185	0.094	[1.997, 2.370]	$\lambda_{102}^{(1)}$	0.745	0.096	[0.568, 0.944]
$\mu_6^{(1)}$	2.321	0.090	[2.138, 2.497]	$\lambda_{112}^{(1)}$	1.066	0.113	[0.857, 1.297]
$\mu_7^{(1)}$	1.931	0.088	[1.760, 2.106]	$\lambda_{122}^{(1)}$	0.900	0.101	[0.712, 1.107]
$\mu_8^{(1)}$	2.055	0.082	[1.897, 2.219]	$\lambda_{143}^{(1)}$	0.716	0.099	[0.529, 0.919]
$\mu_9^{(1)}$	1.459	0.071	[1.319, 1.602]	$\lambda_{153}^{(1)}$	0.892	0.100	[0.703, 1.093]
$\mu_{10}^{(1)}$	1.516	0.071	[1.379, 1.655]	$\lambda_{163}^{(1)}$	0.987	0.096	[0.809, 1.180]
$\mu_{11}^{(1)}$	1.822	0.083	[1.662, 1.984]	$\lambda_{173}^{(1)}$	0.180	0.050	[0.086, 0.279]
$\mu_{12}^{(1)}$	1.473	0.077	[1.321, 1.625]	$\lambda_{183}^{(1)}$	0.703	0.085	[0.543, 0.875]
$\mu_{13}^{(1)}$	1.815	0.102	[1.617, 2.021]	$\phi_{11}^{(1)}$	0.468	0.069	[0.347, 0.611]
$\mu_{14}^{(1)}$	2.011	0.093	[1.826, 2.195]	$\phi_{12}^{(1)}$	0.442	0.066	[0.324, 0.583]
$\mu_{15}^{(1)}$	2.280	0.098	[2.090, 2.472]	$\phi_{22}^{(1)}$	0.615	0.099	[0.443, 0.825]
$\mu_{16}^{(1)}$	1.943	0.098	[1.755, 2.141]	$\gamma_1^{(1)}$	0.573	0.187	[0.207, 0.951]
$\mu_{17}^{(1)}$	1.122	0.044	[1.035, 1.209]	$\gamma_2^{(1)}$	0.539	0.182	[0.209, 0.933]
$\mu_{18}^{(1)}$	1.517	0.081	[1.358, 1.678]	$\gamma_3^{(1)}$	-0.179	0.308	[-0.795, 0.391]
$\lambda_{21}^{(1)}$	1.374	0.122	[1.146, 1.622]	$\gamma_4^{(1)}$	0.038	0.229	[-0.406, 0.491]
$\lambda_{31}^{(1)}$	1.159	0.117	[0.935, 1.399]	$\gamma_5^{(1)}$	0.133	0.174	[-0.180, 0.484]
$\lambda_{41}^{(1)}$	0.812	0.113	[0.596, 1.040]	$\psi_{\varepsilon\delta}^{(1)}$	0.263	0.039	[0.194, 0.349]

Table 8: **Bayesian Estimation of Multiple Group Nonlinear SEM with Ordered Categorical Treated as a Continuous Normal Variables of Second Group under Prior II**

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\mu_1^{(1)}$	1.319	0.061	[1.200, 1.436]	$\lambda_{51}^{(1)}$	1.368	0.137	[1.105, 1.645]
$\mu_2^{(1)}$	2.023	0.079	[1.871, 2.179]	$\lambda_{61}^{(1)}$	1.475	0.129	[1.235, 1.749]
$\mu_3^{(1)}$	1.996	0.079	[1.842, 2.152]	$\lambda_{82}^{(1)}$	1.165	0.120	[0.939, 1.411]
$\mu_4^{(1)}$	1.484	0.067	[1.353, 1.614]	$\lambda_{92}^{(1)}$	0.935	0.096	[0.753, 1.130]
$\mu_5^{(1)}$	2.080	0.086	[1.913, 2.248]	$\lambda_{102}^{(1)}$	0.594	0.081	[0.439, 0.759]
$\mu_6^{(1)}$	2.060	0.081	[1.901, 2.217]	$\lambda_{112}^{(1)}$	1.177	0.109	[0.976, 1.408]
$\mu_7^{(1)}$	1.716	0.078	[1.561, 1.869]	$\lambda_{122}^{(1)}$	0.797	0.083	[0.644, 0.972]
$\mu_8^{(1)}$	2.122	0.085	[1.951, 2.288]	$\lambda_{143}^{(1)}$	0.876	0.089	[0.706, 1.053]
$\mu_9^{(1)}$	1.409	0.070	[1.271, 1.547]	$\lambda_{153}^{(1)}$	0.931	0.084	[0.770, 1.100]
$\mu_{10}^{(1)}$	1.376	0.060	[1.257, 1.494]	$\lambda_{163}^{(1)}$	0.973	0.078	[0.828, 1.134]
$\mu_{11}^{(1)}$	1.622	0.078	[1.475, 1.775]	$\lambda_{173}^{(1)}$	0.276	0.052	[0.175, 0.380]
$\mu_{12}^{(1)}$	1.306	0.060	[1.190, 1.423]	$\lambda_{183}^{(1)}$	0.683	0.070	[0.552, 0.824]
$\mu_{13}^{(1)}$	1.553	0.086	[1.385, 1.723]	$\phi_{11}^{(1)}$	0.414	0.061	[0.306, 0.544]
$\mu_{14}^{(1)}$	1.720	0.092	[1.542, 1.900]	$\phi_{12}^{(1)}$	0.419	0.063	[0.311, 0.556]
$\mu_{15}^{(1)}$	1.786	0.086	[1.620, 1.957]	$\phi_{22}^{(1)}$	0.621	0.099	[0.446, 0.838]
$\mu_{16}^{(1)}$	1.751	0.083	[1.590, 1.916]	$\gamma_1^{(1)}$	1.016	0.130	[0.776, 1.290]
$\mu_{17}^{(1)}$	1.068	0.051	[0.969, 1.167]	$\gamma_2^{(1)}$	0.012	0.063	[-0.114, 0.136]
$\mu_{18}^{(1)}$	1.332	0.071	[1.192, 1.471]	$\gamma_3^{(1)}$	-0.203	0.330	[-0.760, 0.463]
$\lambda_{21}^{(1)}$	1.513	0.125	[1.279, 1.770]	$\gamma_4^{(1)}$	0.234	0.260	[-0.271, 0.746]
$\lambda_{31}^{(1)}$	1.519	0.128	[1.278, 1.784]	$\gamma_5^{(1)}$	0.268	0.190	[-0.104, 0.596]
$\lambda_{41}^{(1)}$	0.921	0.108	[0.718, 1.140]	$\psi_{\varepsilon\delta}^{(1)}$	0.236	0.032	[0.182, 0.306]

Table 9: **Performance of Deviance Information Criterion DIC for Multiple Group SEMs under Prior I**

	Ordered Categorical	Normal
Group 1, n=200	7815.310	9581.190
Group 2, n=200	6755.430	8969.070
Total	14570.700	18550.300

Table 10: **Performance of Deviance Information Criterion DIC for Multiple Group SEMs under Prior II**

	Ordered Categorical	Normal
Group 1, n=200	7820.740	9575.550
Group 2, n=200	6756.540	8964.210
Total	14577.300	18539.800

The estimated multiple group nonlinear structural equation in prior I is given by

$$\eta_i^{(1)} = 0.411\xi_{i1}^{(1)} + 0.788\xi_{i2}^{(1)} + 0.287\xi_{i1}^{(1)}\xi_{i2}^{(1)} - 0.295\xi_{i1}^{(1)}\xi_{i1}^{(1)} - 0.186\xi_{i2}^{(1)}\xi_{i2}^{(1)}, \quad (24)$$

$$\eta_i^{(2)} = 1.178\xi_{i1}^{(2)} + 0.142\xi_{i2}^{(2)} + 0.093\xi_{i1}^{(2)}\xi_{i2}^{(2)} + 0.005\xi_{i1}^{(2)}\xi_{i1}^{(2)} - 0.171\xi_{i2}^{(2)}\xi_{i2}^{(2)}. \quad (25)$$

The estimated multiple group nonlinear structural equation in prior II is given by

$$\eta_i^{(1)} = 0.420\xi_{i1}^{(1)} + 0.782\xi_{i2}^{(1)} + 0.140\xi_{i1}^{(1)}\xi_{i2}^{(1)} - 0.235\xi_{i1}^{(1)}\xi_{i1}^{(1)} - 0.113\xi_{i2}^{(1)}\xi_{i2}^{(1)}, \quad (26)$$

$$\eta_i^{(2)} = 1.194\xi_{i1}^{(2)} + 0.139\xi_{i2}^{(2)} + 0.096\xi_{i1}^{(2)}\xi_{i2}^{(2)} + 0.004\xi_{i1}^{(2)}\xi_{i1}^{(2)} - 0.166\xi_{i2}^{(2)}\xi_{i2}^{(2)}. \quad (27)$$

The results corresponding to the first and second groups under Type I inputs, and ordered categorical variables are reported in Tables (1:2). We observed that the SE values in the first group are smaller than the SE values for the second group. However, it is expected that the empirical performance would be worse with the second group.

The results corresponding to the first and second groups under Type II inputs, and ordered categorical variables are reported in Tables (3:4). We observed that SE values in the first group are smaller than the SE values for the second group. However, it is expected that the empirical performance would be worse with the second group.

The results corresponding to the first and second groups under Type I inputs, and ordered categorical variables when treated as a continuous normal distribution are reported in Table (5:6). We observed that SE values in the first group are smaller than the SE values for the second group. However, it is expected that the empirical performance would be worse with the second group.

The results corresponding to the first and second groups under Type II inputs, and ordered categorical variables when treated as a continuous normal distribution are reported in Tables (7:8). We observed that the SE values in the second group are smaller than the SE values for the first group. However, it is expected that the empirical performance would be worse with the first group. We noted that most of the SE values in ordered categorical variables whan treated as a continuous normal variables are less than the SE values when treated as a ordered categorical variables.

The HPD intervals of all the parameters were computed. We observed that the performances of the HPD intervals are satisfactory for ordered categorical variables when treated as a continuous normal variables.

To reveal the performance of DIC for model comparison, we reanalysed the data sets via a nonlinear structural equation model with interaction term in the structural equation (model 5). The DIC values obtained were compared to those obtained under the correct model. Results are presented in Table 9 and Table 10.

The DIC values are very close in prior I and prior II when using ordered categorical variables. In addition, the DIC values are very close in prior I and prior II when using continuous normal distribution. The model fitting DIC in Group 2 is less than the DIC value in Group 1 in prior I and II and in the ordered categorical variables; when treated as a continuous normal distribution so the model fitting is better in Group 2. As a result, we observed the performance of DIC is not satisfactory and would be worse under ordered categorical variables when trated as a continuous normal distribution However, it performs very well for the ordered categorical variables. However, routinely treating ordered categorical variables as normal may lead to erroneous conclusions (see Lee et al., 1990; Olsson, 1979).

Convergence of the Gibbs sampler are monitored by the plots of several simulated sequences of the individual parameters with different starting values and are presented in Figures 4 and 5 respectively. Bayesian estimates were obtained from T=10000 iterations after discarding 4000 burn-in iterations in linear and nonlinear SEMs.

11. Conclusions and Recommendations

Models involving multiple group nonlinear effects are very common in social and behavioural sciences. The purpose of this analysis was to use multiple group NSEMs to obtain all the estimated parameters and to solve the problem of ordered categorical variables by using hidden continuous normal distribution. However, this assumption is likely to be violated in many practical applications. Developing a non-parametric Bayesian approach to relax the normality assumption in multiple group NSEMs may represent a future research topic.

In SEMs, examples that incorporate nonlinear terms of latent variables in equations exist. As pointed out by Bollen & Paxton,(1998) and Schumacker & Marcoulides, (1998) among others, the lack of applications is not due to the failure of substantive arguments that suggest the presence of nonlinearity, rather the existing statistical methods are

technically demanding and not well understood. In this paper, a Bayesian approach is proposed for analysing multiple group nonlinear models with ordered categorical variables. In addition to point estimation, we provide statistical methods to obtain standard deviations estimates, and model comparison using the Deviance Information Criterion (DIC). Owing to the complexity of the proposed model. As we have seen, difficulties arising from the nonlinear causal relationships among the latent variables and the discrete nature of ordered categorical data manifest variables are alleviated by data augmentation with some MCMC methods. More specifically, the basic idea of our development was inspired by following the common strategy from recent work in statistical computing (see Rubin, 1991) that formulated the underlying complicated problem, so that when augmenting the real observed data with the hypothetical missing data the analysis would be relatively easy with the complete data. This strategy is very powerful and can be applied to other more complex models. The incorrectly treating indicator I as a truncated normal distribution see (Spiegelhalter *et al.*, 2003) and (Lunn et al., 2009).

real example studies is conducted, not only to reveal the empirical performance of the Bayesian approach, but also to show that incorrectly treating ordered categorical as a continuous normal distribution, and vice versa, would produce misleading results.

References

1. Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle.* Paper presented at the Second international symposium on information theory.

2. Bollen, K. A., & Paxton, P. (1998). Interactions of latent variables in structural equation models. *Structural Equation Modeling, 5*, 267-293.

3. Broemeling, L. D. (1985). *Bayesian analysis of linear models*: Dekker New York.

4. Cai, J.-H., Song, X.-Y., & Lee, S.-Y. (2008). Bayesian Analysis of Nonlinear Structural Equation Models with Mixed Continuous, Ordered and Unordered Categorical, and Nonignorable Missing Data. *Statistics and its Interface, 1*, 99-114.

5. Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in medicine*, 28(25), 3049-3067 .

6. Geman, S., & Geman, D. (1984). Stochastic relaxation,Gibbs distribution,and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*(6), 721-741.

7. Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical Science*, 473-483.

8. Greene, W. H. (2003). *Econometric Analysis*: Pearson Education India.

9. Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

10. Lee, S.-Y. and Song, X.-Y. (2012). *Basic and Advanced Structural Equation Models for Medical and Behavioural Sciences*. Hoboken: Wiley.

11. Lee, S.-Y. (2006). Bayesian Analysis of Nonlinear Structural Equation Models with Nonignorable Missing Data. *Psychometrika, 71*(3), 541-564. doi: 10.1007/s11336-006-1177-1

12. Lee, S.-Y. (2007). *Structural Equation Modeling : A Bayesian Approach*. Chichester, England; Hoboken, NJ: Wiley.

13. Lee, S.-Y., Poon, W.-Y., & Bentler, P. (1990). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics & probability letters, 9*(1), 91-97.

14. Lee, S.-Y., & Shi, J.-Q. (2000). Bayesian Analysis of Structural Equation Model With Fixed Covariates. *Structural Equation Modeling: A Multidisciplinary Journal, 7*(3), 411-430. doi: 10.1207/s15328007sem0703_3

15. Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39*(4), 653-686.

16. Lee, S.-Y., Song, X.-Y., & Cai, J.-H. (2010). A Bayesian Approach for Nonlinear Structural Equation Models with Dichotomous Variables Using Logit and Probit Links. *Structural Equation Modeling, 17*(2), 280-302.

17. Lee, S.-Y., Song, X.-Y., & Tang, N.-S. (2007). Bayesian Methods for Analyzing Structural Equation Models With Covariates, Interaction, and Quadratic Latent Variables. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 404-434. doi: 10.1080/10705510701301511

18. Lee, S.-Y., & Tang, N.-S. (2006). Analysis of nonlinear structural equation models with nonignorable missing covariates and ordered categorical data. *Statistica Sinica, 16*(4), 1117.

19. Lee, S.-Y., Poon, W. Y., & Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology, 48*(2), 339-358.

20. Lee, X.-Y. and Song, X.-Y. (2002). A Bayesian Approach for Multigroup Nonlinear Factor Analysis. *Structural Equation Modeling, 9*(4), 523–553.

21. Lee, X.-Y. and Song, X.-Y. (2004). Bayesian Analysis of Two-Level Nonlinear Structural

22. Equation Models with Continuous and Polytomous Data. *British Journal of Mathematical and Statistical Psychology, 57*, 29–52.

23. Lu, B., Song, X.-Y. and Li, X.-D. (2012). Bayesian Analysis of Multi-Group Nonlinear Structural Equation Models with Application to Behavioral Finance. *Quantitative Finance*, 12(3), 477-488. doi: 10.1080/14697680903369500

24. Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*: CRC press.

25. Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in medicine*, 28(25), 3049-3067.

26. Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*(4), 443-460.

27. Rubin, D. B. (1991). EM and beyond. *Psychometrika, 56*(2), 241-254.

28. Schumacker, R. E., & Marcoulides, G. A. (1998). *Interaction and nonlinear effects in structural equation modeling*: Lawrence Erlbaum Associates Publishers.

29. Scott Long, J. (1997). Regression Models for Categorical and Limited Dependent Variables. *Advanced quantitative techniques in the social sciences*, 7.

30. Shi, J.-Q., & Lee, S.-Y. (1998). Bayesian sampling-based approach for factor analysis model with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology, 51*, 233-252.

31. Song, X.-Y., Lu, Z.-H., Hser, Y.-I., & Lee, S.-Y. (2011). A Bayesian Approach for Analyzing Longitudinal Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(2), 183-194. doi: 10.1080/10705511.2011.557331

32. Song, X.-Y., & Lee, S.-Y. (2007). Bayesian analysis of latent variable models with nonignorable missing outcomes from exponential family. *Statistics in medicine, 26*, 681–693.

33. Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). Winbugs user manual. *Medical Research Council Biostatistics Unit, Cambridge*.

34. Spiegelhalter, D. J., Best, N. G., Carlin, B. P.,, & & van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit (with Discussion). *Journal of the Royal Statistical Society, Series B, 64*(4), 583-639.

35. Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398), 528-540.

36. Thanoon, T. Y.& Adnan, R.(2015). Bayesian Analysis of Multiple Group Nonlinear Structural Equation Models with Ordered Categorical and Dichotomous Variables: A Survey. *Research Journal of Mathematical and Statistical Sciences*. Vol. 3(12), 5-15.

37. Thanoon, T. Y.& Adnan, R.(2016a). Bayesian Analysis of Linear and Nonlinear Latent Variable Models with Fixed Covariate and Ordered Categorical Data. *Pakistan Journal of Statistics and Operation Research*, *12*(1), 125-140.

38. Thanoon, T. Y., & Adnan, R. (2016b). Model Comparison of Linear and Nonlinear Bayesian Structural Equation Models with Dichotomous Data. *Communications in Statistics-Simulation and Computation*, 0(0), 1-22. DOI: 10.1080/03610918.2015.1122052.

39.   [38]   Thanoon, T. Y., & Adnan, R. (2016c). Bayesian Multi-Sample Nonlinear Structural Equation Models with Mixed Ordered Categorical and Dichotomous Variables. *Journal of Applied Probability* and Statistics, *11*(2), 1-16.

40.   [39]   Thanoon, Thanoon Y., and Robiah Adnan. (2016d). Comparison between Bayesian structural equation models with ordered categorical data. *AIP Conference Proceedings*. Vol. 1750. No. 1. AIP Publishing.

41.   [40]   Thanoon, T. Y., Adnan, R., & Md Jedi, M. A. B. (2017). Model comparison of Bayesian structural equation models with mixed ordered categorical and dichotomous data. *Journal of Statistics and Management Systems*, 20(1), 113-131. doi:10.1080/09720510.2016.1238111.

42.   [41] Wang, J., & Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus*: John Wiley & Sons.

43.   [42] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*: MIT press.

44.   [43] Yang, M., & Dunson, D. B. (2010). Bayesian Semiparametric Structural Equation Models with Latent Variables. *Psychometrika, 75*(4), 675-693. doi: 10.1007/s11336-010-9174-4.