Bayesian Analysis of Linear and Nonlinear Latent Variable Models with Fixed Covariate and Ordered Categorical Data

Thanoon Y. Thanoon
Department of Mathematical Science, Faculty of Science
Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia
Department of Operation Management Technique, Technical College of Management,
Northern Technical University, Mosul, Iraq
ytthanoon2@live.utm.my

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

Abstract

In this paper, ordered categorical variables are used to compare between linear with covariate and nonlinear interactions of covariates and latent variables in Bayesian structural equation models. Gibbs sampling method is applied for estimation and model comparison. Hidden continuous normal distribution (censored normal distribution) is used to handle the problem of ordered categorical data. Statistical inferences, which involve estimation of parameters and their standard deviations, and residuals analyses for testing the selected model, are discussed. The proposed procedure is illustrated by a real data. Analyses are done by using OpenBUGS program.

Keywords: Structural equation models, Bayesian analysis, latent variables, Gibbs sampling, ordered categorical data.

1.   Introduction

Structural equation modeling (SEM) is a statistical approach to testing hypotheses about the relationships among observed and latent variables. Observed variables also called indicator variables or manifest variables. Latent variables also denoted as unobserved variables or factors. Examples of latent variables in education are math ability and intelligence and in psychology are depression and self confidence. The latent variables cannot be measured directly. Researchers must define the latent variable in terms of observed variables (Khine, 2013).

At present, most statistical theory and computer software in the field of SEMs are based on models that involve only linear relationships among the manifest and the latent variables. More statistically sound methods for linear and nonlinear SEMs and factor analysis have been proposed by Lee and Song (2003), Lee and Song (2005), Lee (2006) **,** Lee and Tang (2006), Cai *et al.* (2008) , Lee *e*t al. (2009), Lee *et al.* (2010), respectively. We used the Bayesian approach to develop methods for statistical inference. MCMC methods, such as the Gibbs Sampler (Geman and Geman, 1984) is used in this paper. Theoretically, in the light of the extension of simple linear regression to multiple regression and nonlinear regression, the importance of generalizing linear 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 linear and 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 is to handle the ordered categorical variables in the Bayesian analysis and 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. Model Description is described in section 2. Bayesian estimation of structural equation models which contain Linear and nonlinear models are described in Section 3. Models comparison using (DIC) are described Section 4. A case study is presented in section 5. Empirical results, which are obtained from a case study, are discussed in Section 6. Some concluding remarks are given in section 7.

2.   Model Description

Consider the following measurement equation for a $p \times 1$ manifest random vector y_i

$$y_i = \mu + \Lambda w_i + \varepsilon_i \quad i = 1,...,\text{n}, \tag{1}$$

where $\mu(p \times 1)$ is the vector of intercepts, $\Lambda(p \times q)$ is the factor loading matrix, $w_i(q \times 1)$ is a latent random vector and $\varepsilon_i(p \times 1)$ is a random vector of error measurements with distribution $N[0, \Psi_\varepsilon], \Psi_\varepsilon$ is diagonal and ε_i is independent with w_i. We let $\eta_i(q_1 \times 1)$ and $\xi_i(q_2 \times 1)$ be latent subvectors of w_i, and consider the following structural equation:

$$\eta_i = \Pi \eta_i + \Gamma F(x_i, \xi_i) + \delta_i \tag{2}$$

Where $\Pi(q_1 \times q_1)$ and $\Gamma(q_1 \times q_2)$ are matrices of regression coefficients, ξ_i and δ_i are independently distributed as $N[0, \Phi]$ and $N[0, \Psi_\delta]$, where Ψ_δ is a diagonal covariance matrix. It is assumed that $\Pi_0 = |I_{q1} - \Pi|$ is a nonzero constant that is independent with elements in Π. $x_i = (x_{i1})^T$ is a vector of covariate.

Let $Y = (y_1,...,y_n)$ is underlying latent continuous measurements are unobservable. The information associated with **y** is given by an observable ordered categorical vector $Z = (z_1,...,z_n)$. That is, any latent variable may have continuous and/or ordered categorical manifest variables as its indicators. The relationship between y and z is defined by a set of thresholds as follows:

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_s \end{bmatrix} \quad if \quad \begin{bmatrix} \alpha_{1,z_1} < y_1 \le \alpha_{1,z_1+1} \\ \vdots \\ \alpha_{s,z_s} < y_1 \le \alpha_{s,z_s+1} \end{bmatrix} \qquad (3)$$

Where for $k = 1,...,s, z_k$ is an integral value in $\{0,1,...,b_k\}$ and $\alpha_{k,0} < \alpha_{k,1} < ... < \alpha_{k,b_k} < \alpha_{k,b_{k+1}}$. In general, we set $\alpha_{k,0} = -\infty, \alpha_{k,b_{k+1}} = \infty$, For the kth variable, there are b_{k+1} categories which are defined by the unknown thresholds $\alpha_{k,j}$. The integral values $\{0,1,...,b_k\}$ of z_k are just used for specifying the categories that contain the corresponding elements in y_k (Lee, 2007).

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 and Lee, 1998) of fixing some thresholds at preassigned values. For convenience, we assume that the positions of the fixed elements are the same for each group.

3. Bayesian estimation of structural Equation models

The objective of this section is to describe a Bayesian approach for analyzing the preceding nonlinear structural equation models in the context of ordered categorical data. Nice features of a Bayesian approach include the following: (a) Prior knowledge can be directly incorporated in the analysis. As a result, more accurate parameter estimates can be obtained under situations with good prior information; (b) As mentioned by many articles on Bayesian analysis of structural equation models (Lee, 2006; Lee and Shi, 2000; Lee *et al.*, 2010; Lee *et al.*, 2007; Song and Lee, 2002, 2004; Song *et al.*, 2011; Yang and Dunson, 2010), the sampling-based Bayesian methods do not rely on asymptotic theory; and (c) The Bayesian estimates and the ML estimates have the same optimal asymptotic properties.

We will utilize the useful strategy of data augmentation described in the Bayesian estimation of SEMs with ordered categorical variables. Let $Z = (z_1,...,z_n)$ be the ordered categorical data matrix, and let $Y = (y_1,...,y_n)$ and $\Omega = (w_1,...,w_n)$ be the matrices of latent continuous measurements and latent variables, respectively. The observed data [Z] is augmented with the latent data $[Y, \Omega]$ in the posterior analysis. The joint Bayesian estimates of Ω. To describe the Bayesian approach for the proposed nonlinear structural equation model, let $Z = \{z_1,...,z_n\}$ be the observed data set of ordered categorical variables, and θ be vector that contains the unknown free parameters under the model M.

In a Bayesian approach, θ is considered random with a prior distribution and a prior density function, say $p(\theta)$. Bayesian inference is based on the observed data Z and $p(\theta)$.

Let $p(Z,\theta)$ be the joint probability density function of Z and θ under model M. Based on a basics in probability, $p(Z,\theta) = p(Z|\theta)p(\theta)$, where $p(Z|\theta)$ and $p(\theta|Z)$ are conditional density functions; it follows that

$$\log p(\theta|Z) \propto \log p(Z|\theta) + \log p(\theta) \tag{4}$$

The function $p(\theta|Z)$ is called the posterior density function of the unknown parameters. In Equation (4) the posterior density function $p(\theta|Z)$ depends on likelihood function $p(Z|\theta)$ and prior distribution $p(\theta)$. The function $p(Z|\theta)$ depends on the sample size, whereas $p(\theta)$ does not. For large samples, the prior of θ plays a less important role, and the posterior density function $p(\theta|Z)$ is close to the likelihood function $p(\theta)$. Thus, the Bayesian and ML approaches are asymptotically equivalent, and the Bayesian estimates have the same optimal asymptotical properties as the ML estimates. However, $p(\theta)$ plays a significant role in the Bayesian approach in the situations where the sample size is small, or the information given by ordered categorical data Z.

In this article, we define the Bayesian estimate of θ as the mean of the posterior distribution (called the posterior mean). For simple structural equation models, the posterior mean can be obtained through direct integration. However, due to the complexity of the proposed nonlinear structural equation model with covariates and dichotomous variables, the relating integral does not have a closed form. We apply some MCMC methods in statistical computing to solve this problem. Let y_i be the unobserved variables that correspond to the manifest ordinal variables in z_i, and let $Y = (y_1,...,y_n)$ and $\Omega = (w_1,...,w_n)$. If we can draw a sufficiently large number of observations $\{(\theta^{(t)},\Omega^{(t)},Y^{(t)}); t = 1,...,T\}$ from the joint posterior distribution $p(\theta, \Omega, Y|Z)$, then the Bayesian estimate of θ and the standard error estimates can be obtained from the following sample mean and sample variance matrix, respectively:

$$\widehat{\theta} = T^{-1}\sum_{t=1}^{T}\theta^{(t)}, \quad \widehat{\mathrm{var}}(\theta|Z) = (T-1)^{-1}\sum_{t=1}^{T}(\theta^{(t)} - \widehat{\theta})(\theta^{(t)} - \widehat{\theta})'. \tag{5}$$

It is necessary to specify the prior distributions for components in θ when deriving the conditional distribution of θ given (Ω, Y, Z) in Step a. In general Bayesian analyses, the conjugate prior distributions have been found to be flexible and convenient (Broemeling, 1985).

This kind of prior distribution has been widely applied to many Bayesian analyses in structural equation models (Lee and Song, 2004 ; Song and Lee, 2007). Hence, the following well-known conjugate prior distributions are used:

$$p(\mu) \approx N[\mu_0, H_{0\mu}], \quad p(\lambda_k) \approx N[\lambda_{0k}, H_{0k}], \quad p(\lambda_{\xi k}|\psi_{\delta k}) \approx N[\lambda_{0\xi k}, \psi_{\delta k}H_{0\xi k}],$$
$$p(\Phi^{-1}) \approx W_q[R_0, \rho_0], \quad p(\psi_{\delta k}^{-1}) \approx Gamma[\alpha_{0k}, \beta_{0k}] \tag{6}$$

Where $\psi_{\delta k}$ is the *k*th diagonal element of Ψ_{δ}, λ_k' and $\lambda_{\xi k}'$ are the *k*th rows of Λ and Λ_{ξ}, respectively. $H_{0\mu} = diag(\sigma_{01}^2,...,\sigma_{0p}^2)$, 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 given by the prior information. In general, prior information can be obtained from causal observation or theoretical consideration of experts, or the analyses of past data. As pointed out by Kass and Raftery (1995), priors are often picked for convenience when there is a lack of accurate prior knowledge, because the effect of the priors in Bayesian estimation is small when the sample size is fairly large. For completeness, the conditional posterior distributions of Ω, Y and the components of θ based on the conjugate prior distributions. These results are useful for writing the computer program to implement the Gibbs sampler.

In the Bayesian approach, we need to evaluate the posterior distribution $[\alpha, \theta, \Omega | Z]$. This distribution is rather complicated. To capture its characteristics, we will try to draw a sufficiently large number of observations from it such that the empirical distribution of the generated observations is a close approximation to the true distribution. A good candidate for simulating observations from the posterior distribution is the Gibbs sampler (Geman and Geman, 1984), which iteratively simulates α, θ and Ω from the full conditional distributions. However, owing to the presence of the ordered categorical variables, these conditional distributions are rather complicated to derive and simulating observations from them is difficult. This motivates the further augmentation of the latent matrix *Y* in the posterior analysis, and the consideration of the joint posterior distribution $[\alpha, \theta, \Omega, Y | Z]$. To implement the Gibbs sampler for generating observations of this posterior distribution, we start with initial starting values $(\alpha^{(0)}, \theta^{(0)}, \Omega^{(0)}, Y^{(0)})$, then simulate $(\alpha^{(1)}, \theta^{(1)}, \Omega^{(1)}, Y^{(1)})$ and so on according to the following procedure. At the mth iteration with current values $\alpha^{(m)}, \theta^{(m)}, \Omega^{(m)}, Y^{(m)}$

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

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 $p(\alpha, Y | \theta^{(m+1)}, \Omega^{(m+1)}, Z)$ (7)

The cycle defined above generates $p(\alpha, Y | \theta^{(m+1)}, \Omega^{(m+1)}, Z)$ after the *m*th iteration. As *m* approaches infinity, the joint distribution of $(\alpha^{(m)}, \theta^{(m)}, \Omega^{(m)}, Y^{(m)})$ can be shown to approach the joint posterior distribution $[\alpha, \theta, \Omega, Y | Z]$. (see Geman and Geman, (1984); Geyer, (1992)).

The sequences of the quantities simulated from the joint posterior distribution will be used to calculate the Bayesian estimates and other related statistics. Convergence of the Gibbs sampler can be monitored by the plots of several simulated sequences of the individual parameters with different starting values. The sequences of the quantities

simulated from the joint posterior distribution will be used to calculate the Bayesian estimates and other related statistics.

4. Model Comparisons

A model comparison statistic that takes into account the number of unknown parameters in the model is the DIC (see Spiegelhalter *et al.*, 2002). This statistic is intended as a generalization of the Akaike Information Criterion (AIC; Akaike, (1973) . Under a competing model M_k with a vector of unknown parameters θ_k of dimension d_k, let $\{\theta_k^{(t)} : t = 1,...,T \}$ be a sample of observations simulated from the posterior distribution. The DIC for M_k is computed as follows:

$$DIC_k = -\frac{2}{T}\sum_{t=1}^{T}\log p(Z \mid \theta_k^{(t)}, M_k) + 2d_k, \tag{8}$$

In model comparison, the model with the smaller DIC value is selected. As mentioned in Spiegehalter *et al.* (2003), practical applications of DIC, it is important to note the followings:

(a) If the difference in DIC is small, for example less than 5, and the models make very different inferences, then just reporting the model with the lowest DIC could be misleading. (b) DIC can be applied to non-nested models. (c) Similar to the Bayes factor (Kass and Raftery, 1995) BIC, and AIC, DIC gives a clear conclusion to support the null hypothesis or the alternative hypothesis.

To illustrate the use of DIC for model comparison, we analyzed the same data by a linear and a nonlinear structural equation model with the same measurement model:

$$\eta_i = \beta_1 * \mathrm{x}(\mathrm{i},1) + \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \delta_i, \tag{9}$$

$$\eta_i = \beta_1 * \mathrm{x}(\mathrm{i},1) + \beta_2 * \mathrm{x}(\mathrm{i},1) * \xi_{i2} + \beta_3 * \mathrm{x}(\mathrm{i},1) * \xi_{i3} + \beta_4 * \mathrm{x}(\mathrm{i},1) * \xi_{i2}\xi_{i2} + \beta_5 * \mathrm{x}(\mathrm{i},1) * \xi_{i1}\xi_{i2} + \gamma_1\xi_{i1}$$
$$+ \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \gamma_4\xi_{i2}\xi_{i2} + \delta_i \tag{10}$$

The DIC value corresponding to the linear and nonlinear structural equation model produced by OpenBUGS. The correct nonlinear structural equation model is selected.

5. Case Study and Example

The quality of life data set was established by (Power et al,. 1999) to evaluate three latent variables $(\eta, \xi_{i1}, \xi_{i2}, \xi_{i3})$. Some of these variables are selected in this paper. The first three items (Q3 to Q5) are intended to address physical health, the next three items (Q6 to Q8) are intended to address psychological health, the three items (Q9, Q10, Q11) that follow are for social relationships, and the last one items (Q12) are intended to address environment. The instrument also includes two ordered categorical items for the overall QOL (Q1) and general health (Q2), giving a total of 11 items. All of the items are measured with a five-point scale (1 = 'not at all/very dissatisfied'; 2 = 'a little/dissatisfied'; 3 = 'moderate/neither'; 4 = 'very much/satisfied'; 5 = 'extremely/very

satisfied'). The sample size of the whole data set is extremely large. To illustrate the Bayesian methods, we only analyze a synthetic data set with sample size n = 200.

Our Bayesian linear & nonlinear SEMs defined in Equation (9) and Equation (10) respectively. Hence, some quadratic and interaction effects of the latent variables are considered. To illustrate the Bayesian methods in analyzing linear and nonlinear structural equation models with ordered categorical variables, we use a real data set that is related to random vectors $z_i = (z_{1i}, z_{2i}, ..., z_{11i})'$, let $y_i = (y_{1i}, y_{2i}, ..., y_{11i})'$ be the latent continuous random vector corresponds to the ordinal variables $z_{1i}, z_{2i}, ..., z_{11i}$ where $z_i, i = 1, ..., n$ are ordered categorical variables that are related with 4 latent variables $w_i = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3})$, $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{i11})$, with the following values of the parameters in $\mu = (\mu_1, \mu_2, ..., \mu_{11})'$ and $\Lambda = (\lambda_1', \lambda_2', ..., \lambda_{11}')'$

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{bmatrix} \qquad (11)$$

The relationships of the latent variables in $w_i = (\eta_i, \xi_{1i}, \xi_{2i})$ are assessed by the nonlinear structural equation which is described in Equation (10). A covariate x (200*1) and those corresponding to Φ_1, Φ_2 and Φ_3 are $\rho_0 = 30, R_0^{-1} = 8$, respectively. 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 Equation (6) are set equal to the following values and initial values are equal to 1;

Prior I. Elements in μ_0, λ_{0k} and $\lambda_{0\xi k}$ in Equation (6) are set equal to the true values;

$R_0^{-1} = 8\Phi$, 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.

The parameter values were analyzed by OpenBUGS. In checking convergence, we observed that most parameters such as $\mu, \lambda, \gamma, \Phi, \psi_{\varepsilon\delta}$ converged quickly see Figures 3 and 4. Comparing the Bayesian analyses of structural equation models with data, the MCMC procedure for analyzing data required more iterations to converge. Bayesian estimates were obtained from T=10000 iterations after discarding 5000 burn-in iterations. The Bayesian estimates of nonlinear SEM with covariate and the 95% HPD intervals are presented in Table 1. The Bayesian estimates of linear SEM with covariate and the 95% HPD intervals are presented in Table 2. Estimates of the latent variables were also been obtained from OpenBUGS. The performances of deviance information criterion when comparing linear and nonlinear models are presented in Table 3.

6. Results and Discussion

The objective of this section is to present results of a case 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:

$$Model\ 1: \eta_i = \beta_1 * x(i,1) + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \delta_i,$$

$$Model\ 2: \eta_i = \beta_1 * x(i,1) + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \gamma_4 \xi_{i1} \xi_{i2} + \delta_i,$$

$$Model\ 3: \eta_i = \beta_1 * x(i,1) + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \gamma_4 \xi_{i1} \xi_{i3} + \delta_i,$$

$$Model\ 4: \eta_i = \beta_1 * x(i,1) + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \gamma_4 \xi_{i2} \xi_{i3} + \delta_i,$$

$$Model\ 5: \eta_i = \beta_1 * x(i,1) + \beta_2 * x(i,1) * \xi_{i2} + \beta_3 * x(i,1) * \xi_{i3} + \beta_4 * x(i,1) * \xi_{i2} \xi_{i2} +$$

$$\beta_5 * x(i,1) * \xi_{i1} \xi_{i2} + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \gamma_4 \xi_{i2} \xi_{i2} + \delta_i \qquad (12)$$

In this paper, a Bayesian approach is introduced for analysing linear and nonlinear SEMs with ordered categorical variables. 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 OpenBUGS. Therefore, our proposed method can be conveniently applied on real data. The purpose of this analysis is to compare between linear and nonlinear Bayesian 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 behavioral, educational, medical and social sciences, data are often in ordered categorical variables with observations in discrete form. In analyzing ordered categorical data, the basic assumption in SEM is that the data comes from a continuous normal distribution which 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 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 to relax the normality assumption in NSEMs may represent a future research topic.

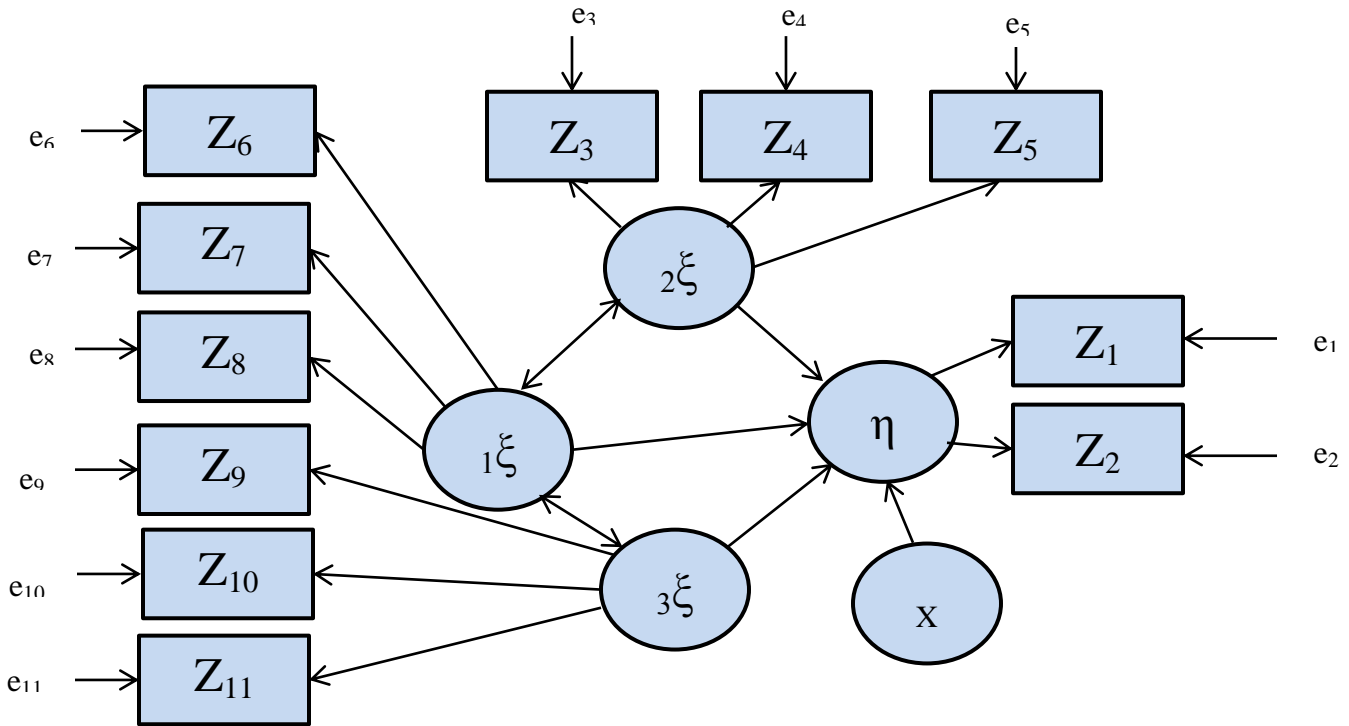**Figure 1**:   The path diagram of linear SEM (model 1)
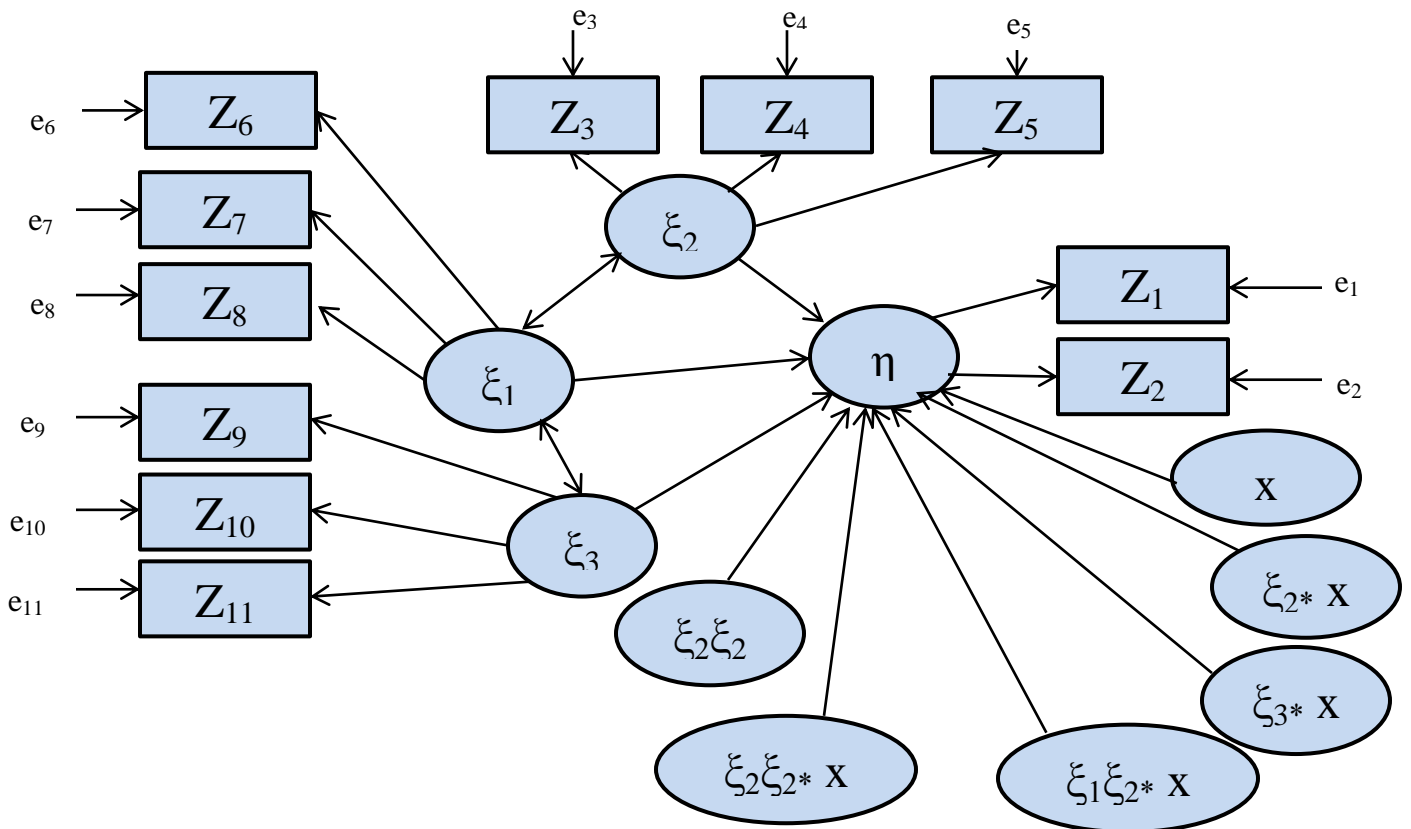
Figure 2. The path diagram of nonlinear SEM (model 5).
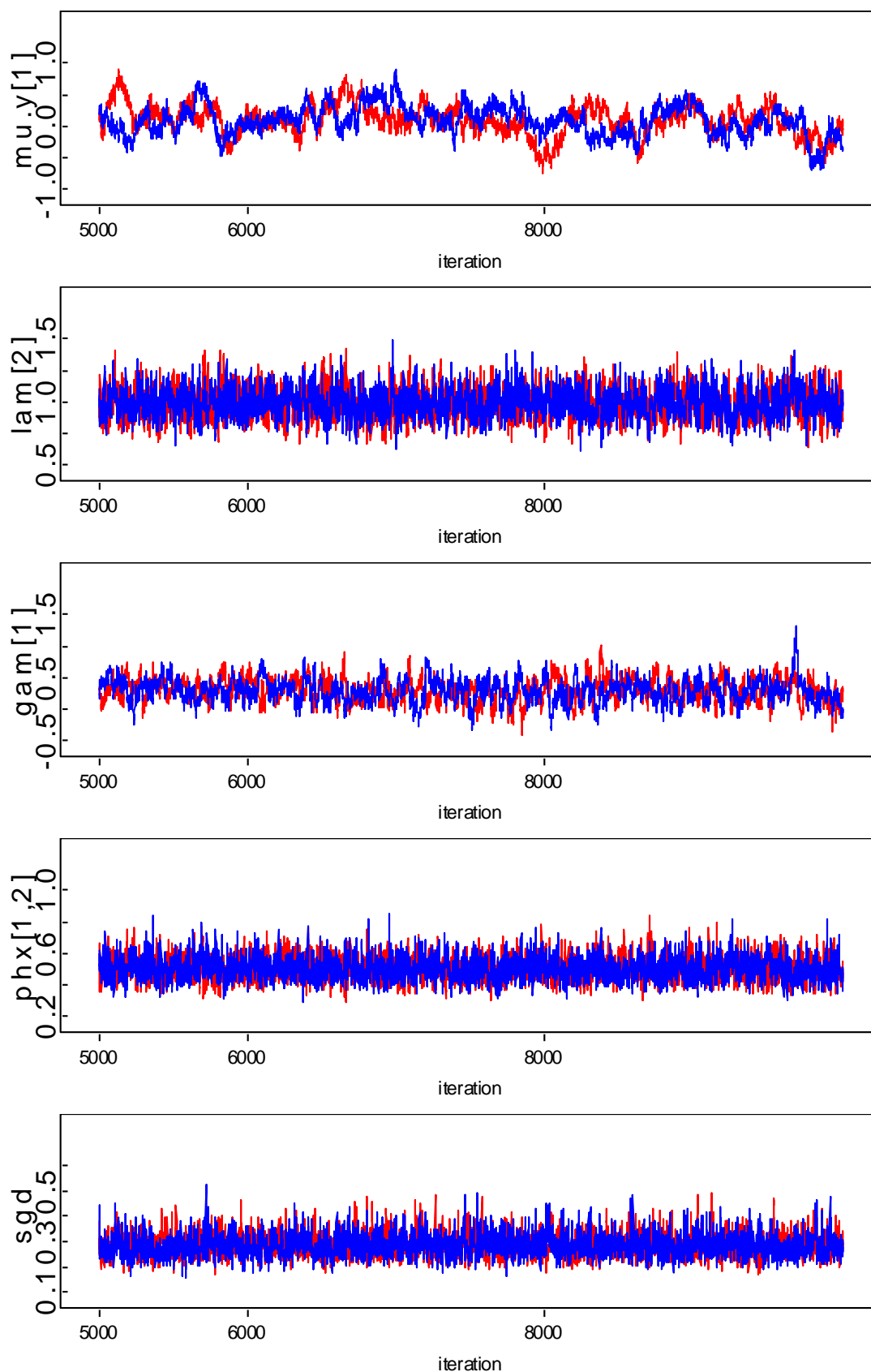
Figure 3. Sequences of (a) μ_1; (b) λ_2; (c) γ_1; (d) Φ_{12}; and (e) $\psi_{\varepsilon\delta}$ for linear SEM
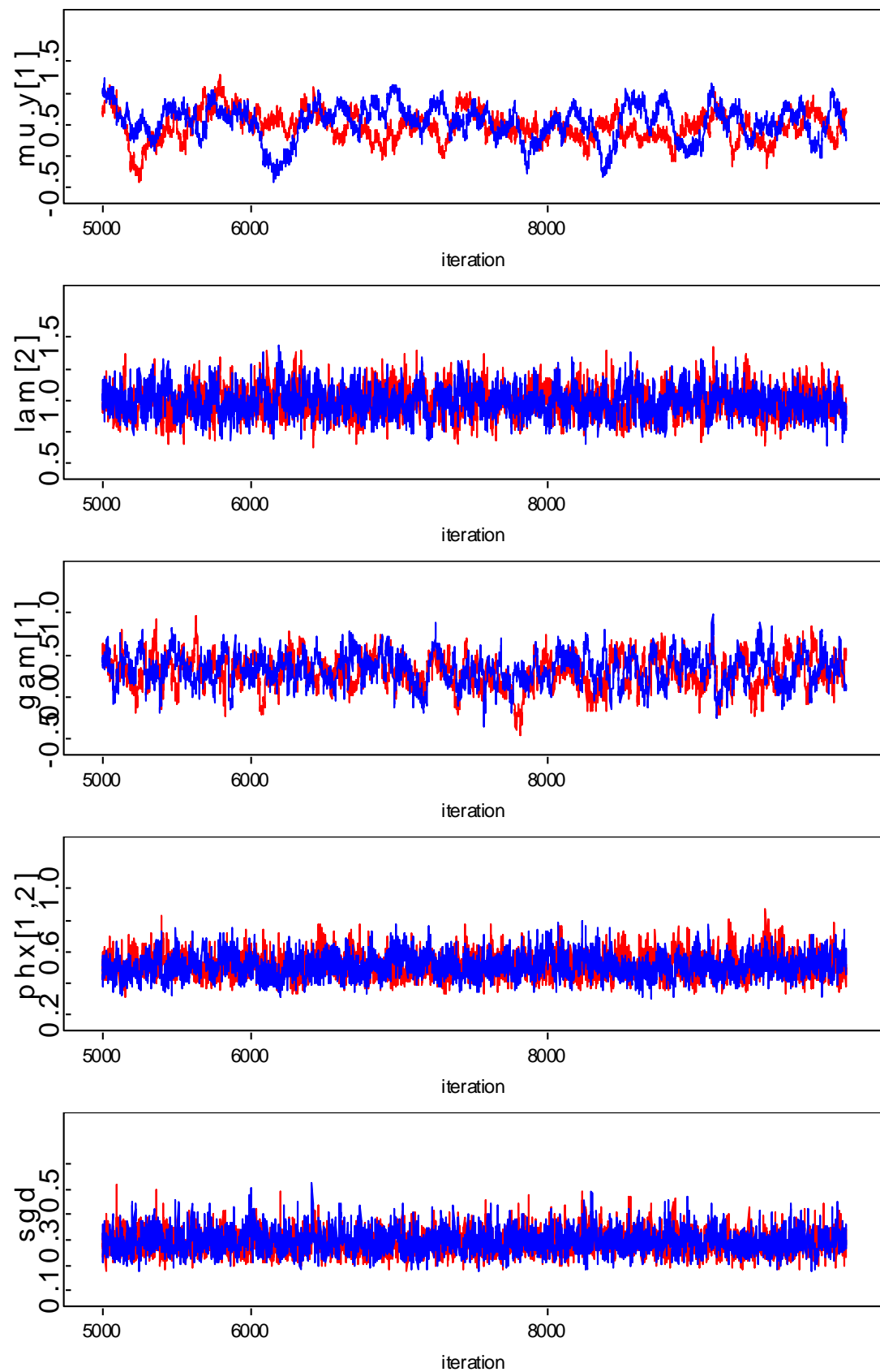
Figure 4. Sequences of (a) μ_1; (b) λ_2; (c) γ_1; (d) Φ_{12}; and (e) $\psi_{\varepsilon\delta}$ for nonlinear SEM

Table 1. Bayesian Estimation of nonlinear SEM with Ordered Categorical Variables

Para	Est.	SD	HPD Interval	Para	Est.	SD	HPD Interval
μ_1	0.504	0.261	[-0.083,0.975]	ϕ_{12}	0.510	0.072	[0.382,0.669]
μ_2	0.161	0.222	[-0.318,0.555]	ϕ_{13}	0.327	0.089	[0.174,0.518]
μ_3	-0.097	0.072	[-0.236,0.044]	ϕ_{21}	0.510	0.072	[0.382,0.669]
μ_4	0.089	0.081	[-0.065,0.253]	ϕ_{22}	0.683	0.091	[0.527,0.880]
μ_5	0.074	0.073	[-0.071,0.219]	ϕ_{23}	0.374	0.098	[0.203,0.584]
μ_6	0.067	0.081	[-0.090,0.229]	ϕ_{31}	0.327	0.089	[0.174,0.518]
μ_7	-0.021	0.075	[-0.168,0.125]	ϕ_{32}	0.374	0.098	[0.203,0.584]
μ_8	0.099	0.080	[-0.060,0.256]	ϕ_{33}	0.477	0.146	[0.249,0.808]
μ_9	0.101	0.076	[-0.052,0.249]	γ_1	0.314	0.190	[-0.072, 0.665]
μ_{10}	-0.298	0.043	[-0.384,-0.212]	γ_2	0.449	0.362	[-0.270, 1.153]
μ_{11}	-0.216	0.047	[-0.307,-0.123]	γ_3	0.416	0.408	[-0.320, 1.259]
λ_1	0.836	0.081	[0.689,1.005]	γ_4	0.067	0.277	[-0.433, 0.633]
λ_2	0.988	0.113	[0.783,1.226]	β_1	-0.076	0.064	[-0.191, 0.064]
λ_3	0.884	0.105	[0.687,1.104]	β_2	0.060	0.083	[-0.101, 0.224]
λ_4	0.718	0.087	[0.555,0.895]	β_3	-0.042	0.101	[-0.245, 0.152]
λ_5	1.086	0.081	[0.931,1.250]	β_4	-0.006	0.076	[-0.176, 0.135]
λ_6	0.452	0.097	[0.287,0.664]	β_5	0.001	0.069	[-0.133, 0.144]
λ_7	0.297	0.103	[0.123,0.530]	$\psi_{\varepsilon\delta}$	0.292	0.046	[0.215, 0.392]
ϕ_{11}	0.582	0.089	[0.427,0.773]				

136

Pak.j.stat.oper.res. Vol.XII No.1 2016 pp125-140

Table 2. Bayesian Estimation of linear SEM with Ordered Categorical Variables

Para	Est.	SD	HPD Interval	Para	Est.	SD	HPD Interval
μ_1	0.090	0.238	[-0.416, 0.553]	λ_6	0.452	0.097	[0.277, 0.657]
μ_2	-0.184	0.203	[-0.611, 0.208]	λ_7	0.300	0.101	[0.126, 0.521]
μ_3	-0.180	0.072	[-0.319, -0.036]	ϕ_{11}	0.560	0.088	[0.407, 0.755]
μ_4	0.006	0.079	[-0.145, 0.165]	ϕ_{12}	0.502	0.072	[0.374, 0.658]
μ_5	-0.001	0.073	[-0.142, 0.144]	ϕ_{13}	0.315	0.091	[0.171, 0.525]
μ_6	-0.023	0.077	[-0.172, 0.130]	ϕ_{21}	0.502	0.072	[0.374, 0.658]
μ_7	-0.094	0.071	[-0.234, 0.045]	ϕ_{22}	0.690	0.092	[0.528, 0.889]
μ_8	0.001	0.075	[-0.146, 0.150]	ϕ_{23}	0.370	0.103	[0.209, 0.610]
μ_9	0.010	0.071	[-0.131, 0.147]	ϕ_{31}	0.315	0.091	[0.171, 0.525]
μ_{10}	-0.328	0.042	[-0.411, -0.245]	ϕ_{32}	0.370	0.103	[0.209, 0.610]
μ_{11}	-0.239	0.046	[-0.327, -0.149]	ϕ_{33}	0.468	0.160	[0.255, 0.857]
λ_1	0.853	0.080	[0.703,1.012]	γ_1	0.298	0.186	[-0.065, 0.658]
λ_2	0.992	0.117	[0.776,1.231]	γ_2	0.631	0.177	[0.292, 0.986]
λ_3	0.890	0.106	[0.691,1.108]	γ_3	0.261	0.170	[-0.067, 0.599]
λ_4	0.711	0.085	[0.552,0.883]	β_1	0.007	0.058	[-0.104, 0.129]
λ_5	1.071	0.084	[0.916,1.242]	$\psi_{\varepsilon\delta}$	0.287	0.045	[0.211,0.386]

Table 3. The DIC Values for Linear and Nonlinear EMs with Ordered Categorical Variables and Covariates

Linear with covariates	Nonlinear with covariates
5011.0	5034.0

The results corresponding to the nonlinear SEM with covariates under Type I and II inputs, and ordered categorical variables are reported in Table (1). We observed that the SD values are very small in the nonlinear SEM.

The results corresponding to the linear SEM and covariate under Type I and II inputs, and ordered categorical variables are reported in Table (2). We observed that the SD values are very small in the linear SEM with covariate.

The results corresponding to the linear and nonlinear SEM with covariates under Type I and II inputs, we observed that the SD values in linear SEM are smaller than the SD values for the nonlinear SEM with covariates. However, it is expected that the empirical performance would be worse with the nonlinear SEM with covariates.

The HPD intervals of all the parameters were computed. We observed that the performances of the HPD intervals in linear SEM and covariate are satisfactory for ordered categorical variables. To reveal the performance of DIC for model comparison, we reanalysed the data sets via a nonlinear SEM and covariate in the structural equation (Model 5). The DIC values obtained were compared to those obtained under the correct model. Results are presented in Table 3.

The model fitting DIC in linear SEM is less than the DIC value in nonlinear SEM in the ordered categorical variables. As a result, we observed the performance of DIC is not satisfactory and would be worse under nonlinear effect and ordered categorical variables. However, it performs very well with linear effect with covariate and ordered categorical variables.

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 3 and 4 respectively. Bayesian estimates were obtained from T=10000 iterations after discarding 5000 burn-in iterations in linear and nonlinear SEMs with covariates.

7. Conclusions and Recommendations

The Bayesian linear and nonlinear SEMs which involving covariates are very common in social and behavioural sciences. However, in SEM, examples that incorporate nonlinear and covariate terms of latent variables into structural equations exist. As pointed out by Bollen and Paxton (1998), Schumacker and 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 a linear and nonlinear, covariate models with ordered categorical variables. In addition to point estimation, we provide statistical methods to obtain standard deviation estimates, and model comparisons using the deviance information criterion (DIC). Owing to the complexity of the proposed model, as we have seen, 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 is inspired by the following common strategy from recent work in statistical computing (see Rubin, 1991) that formulate 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.

References

1. Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. Proceedings of the 1973 *Second international symposium on information theory*, 267-281.

2. Broemeling, L. D. (1985). *Bayesian Analysis of Linear Models*: Dekker New York.

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

4. Cai, J.-H., Song, X.-Y. and 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. Geman, S. and Geman, D. (1984). Stochastic Relaxation,Gibbs Distribution,and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6), 721-741.

6. Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science*, 473-483 .

7. Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90, 773–795.

8. Khine, M. S. (2013). Application of Structural Equation Modeling in Educational Research and Practice: Springer.

9. Lee, S.-Y., Poon, W.-Y. and 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.

10. Lee, S.-Y. and Song, X.-Y. (2005). Maximum Likelihood Analysis of a Two-Level Nonlinear Structural Equation Model with Fixed Covariates. *Journal of Educational and Behavioral Statistics*, 30(1), 1-26. doi: 10.3102/10769986030001026

11. Lee, S.-Y., Song, X.-Y., Cai, J. H., So, W. Y., Ma, C. W. and Chan, C. N. (2009). Nonlinear Structural Equation Models with Correlated Continuous and Discrete Data. *Br J Math Stat Psychol*, 62(Pt 2), 327-347. doi: 10.1348/000711008X292343

12. Lee, S.-Y. and Song, X.-Y. (2003). Model Comparison of Nonlinear Structural Equation Models with Fixed Covariates. *PSYCHOMETRIK*, 68(1), 27-47.

13. Lee, S.-Y. and 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.

14. Lee, S.-Y., Song, X.-Y. and 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.

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

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

17. Lee, S.-Y., Song, X.-Y. and 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.

18. Lee, S. and Tang, N. (2006). Analysis of Nonlinear Structural Equation Models with Nonignorable Missing Covariates and ordered categorical data. *Statistica Sinica*, 16(4), 1117.

19. 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

20. 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

21. Lee, S.-Y. and 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

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

23. Lee, X.-Y. S .a. S.-Y. (2004). Bayesian Analysis of Two-Level Nonlinear Structural Equation Models With Continuous and Polytomous Data. *British Journal of Mathematical and Statistical Psychology*, 57, 29–52 .

24. Lee, S.-Y., Poon, W.-Y. and Bentler, P. (1990). Full Maximum Likelihood Analysis of Structural Equation Models with Polytomous Variables. *Statistics & probability letters*, 9(1), 91-97.

25. Power, M., Bullingen, M. and Hazper, A. (1999) The World Health Organization WHOQOL-100: Tests of the universality of quality of life in 15 different cultural groups worldwide. *Health Psychology*, 18, 495–505.

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

27. Schumacker, R. E. and Marcoulides, G. A. (1998). *Interaction and Nonlinear Effects in Structural Equation Modeling*: Lawrence Erlbaum Associates Publishers.

28. 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.

29. Song, X.-Y., Lu, Z.-H., Hser, Y.-I. and 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.

30. Shi, J.-Q. and 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. 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 .

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