

Bayesian Nonlinear Latent variable Models with Mixed Non-normal Variables and Covariates for Multi-sample Psychological Data

Thanoon Y. Thanoon

Department of Business Management Techniques, Administrative Technical College

Northern Technical University, Iraq

Thanoon.younis@ntu.edu.iq

Athar Talal Hamed

Department of Financial and Banking Sciences, College of Administration and Economics

a_talal2013@yahoo.com

Robiah Adnan

Department of Mathematical Sciences, Faculty of Science

University Technology Malaysia, Malaysia

robiah@utm.my

Abstract

The purpose of this paper is to develop a latent variable model with nonlinear covariates and latent variables. Mixed ordered categorical and dichotomous variables and covariates with two different types of thresholds (with equal and unequal spaces) are used in Bayesian multi-sample nonlinear latent variable models and the Gibbs sampling method is applied for estimation and model comparison. Hidden continuous normal distribution (censored normal distribution) and (truncated normal distribution with known parameters) are used to handle the problem of mixed ordered categorical and dichotomous data. Hidden continuous normal distribution (truncated normal distribution with known parameters) is used to handle the problem of mixed ordered categorical and dichotomous data in covariates. Statistical analysis, which involves the estimation of parameters, standard deviations and their highest posterior density, are discussed. The proposed procedure is illustrated using psychological data with the results obtained from the OpenBUGS program.

Keywords: Latent variable Models, Bayesian Analysis, Multi-Sample Data, Mixed Variables, Covariates.

1. Introduction

Latent variable models (LVMs) (Lee, 2007) are a statistical technique for modelling a sequence of correlated data to estimate the interrelationships among manifest and latent variables.

In recent years, many researchers have proposed models that contain nonlinear terms among the manifest and latent variables. Some of these papers have been proposed by Lee and Song (2003), Lee (2006), Cai et al. (2008), Lee et al., (2010), Thanoon and Adnan (2015); Thanoon and Adnan (2016a), Thanoon and Adnan (2016b), Thanoon and Adnan (2016c), Thanoon et al., (2017), Thanoon and Adnan (2017), Thanoon and Adnan (2019).

Lee and Song (2002) offer a specific method for applying the Bayesian approach in factor analysis. They created an analysis model that implements joint Bayesian estimates for the factor scores and structural parameters as they are related to the determined constraints allowing multiple findings to be determined simultaneously. This system, which uses the

Metropolis- Hastings algorithm in combination with the Gibbs model, has been proven efficient in generating computations of these estimates.

Song and Lee (2006b) developed a Bayesian nonlinear latent variable model with nonlinear fixed covariates and latent variables in the structural model and linear fixed covariate and latent variables in the measurement model. Mixed continuous and dichotomous data are used in this study and hidden continuous normal distribution (truncated normal with unknown parameters) is presented to solve the problem of dichotomous data.

Lee (2007) used underlying latent continuous normal distribution (truncated normal distribution with unknown parameters) to solve the problem of ordered categorical variables in Bayesian multi-sample nonlinear latent variable models, and Gibbs sampling method is used to estimate the parameters.

The strategy of multi-sample analysis is very important in many applications, for instance, cross-cultural research. Practically, nonlinear effects, such as quadratic and interaction effects among the covariates and latent variables are important in establishing the substantive theory in many areas. The rapid growth of LVMs is due to the demand of subtle models and the related statistical methods for solving complex research problems in various fields.

The Bayesian approach is developed with the Gibbs sampler algorithm (Geman and Geman, 1984), and the hidden continuous normal measurements and the latent variables in multi groups are treated as hypothetical missing data. Non-informative priors are used for the thresholds (cut points with equal and unequal distances) and conjugate priors are used for the structural parameters.

The main objective of this paper is to propose a Bayesian approach for analysing multi-sample nonlinear LVMs with mixed variables and covariates. The Deviance Information Criterion (DIC; see Spiegelhalter et al., 2002) is used for model comparison.

The main idea is to handle the mixed variables and covariates in the Bayesian analysis and to treat the hidden continuous measurements as a 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 multi-sample latent variable models that contain nonlinear models is described in Section 3. The comparison of models using DIC is described in Section 4. A case study of psychological data is presented in section 5. The results and discussion are described in section 6, and some concluding remarks are given in section 7.

2. Model Description

The latent variable models which have been suggested for application in this case have linear covariates and latent variables within the measurement equation. It also has nonlinear latent variables in the structural model, as well as nonlinear covariates. For example, the following LVM is considered.

$$v_i^{(g)} = A^{(g)}c_i^{(g)} + \Lambda^{(g)}\omega_i^{(g)} + \varepsilon_i^{(g)}, i = 1, \dots, n \quad (1)$$

where $v_i^{(g)} = \{y^{(g)}, u^{(g)}\}$ has been established as a $p \times 1$ random vector of manifest variables, where $y = (y_1, y_2, \dots, y_s)$ is a subset of variables whose exact continuous measurements are unobservable, $u = (u_1, u_2, \dots, u_r)$ is the remaining subset of variables such that $p \geq s = p - r \geq 0$ and the corresponding continuous measurements are unobservable, $A^{(g)}(p \times m_1)$ is a vector of linear covariates, $c_i^{(g)}(m_1 \times 1)$ is a vector of

mixed ordered categorical and dichotomous covariates, $\Lambda^{(g)}(p \times q)$ is thus a matrix with unknown parameters, which is usually considered the factor loading matrix, $\omega_i^{(g)}$ is a $q \times 1$ random vector of latent variables, $\varepsilon_i^{(g)}$ is a $p \times 1$ random vector of residuals. As a result, it is assumed that the resulting $y_i^{(g)}$'s is independent, and that $\omega_i^{(g)}$ is then distributed independently as $N[0, \Phi^{(g)}]$, Further, $\varepsilon_i^{(g)}$ is independently distributed as $N[0, \Psi_\varepsilon^{(g)}]$, where Ψ_ε is a diagonal matrix with diagonal elements $\psi_{\varepsilon_1}, \dots, \psi_{\varepsilon_p}$. It has also been determined that, in this case, $\varepsilon_i^{(g)}$ and $\omega_i^{(g)}$ are both independent as well.

To carry out more complicated mathematical scenarios, ω_i , a latent vector, is subdivided into $(\eta_i^T, \xi_i^T)^T$, in which $\eta_i(q_1 \times 1)$ and $\xi_i(q_2 \times 1)$ are both vectors.

More specifically, $\eta_i(q_1 \times 1)$ is the vector of the endogenous latent variables and $\xi_i(q_2 \times 1)$ is the vector of the exogenous latent variables. In order to create an estimation of the potential significant causal effect of $x_i(m_2 \times 1)$, the vector of mixed ordered categorical and dichotomous covariates on η_i , however, if x_i is non-normal, then ξ_i is non-normal also. Thus, it is important to solve the problem of mixed ordered categorical and dichotomous data in covariates.

The latent variable is defined by the following general latent variable:

$$\eta_i^{(g)} = B^{(g)}\eta_i^{(g)} + \Gamma^{(g)}F(x_i^{(g)}, \xi_i^{(g)}) \quad i = 1, \dots, n \quad (2)$$

where

$B^{(g)}(q_1 \times q_1)$ is a matrix of unknown parameters, $F(x_i^{(g)}, \xi_i^{(g)}) = (f_1(x_i^{(g)}, \xi_i^{(g)}), \dots, f_r(x_i^{(g)}, \xi_i^{(g)}))^T$ is a vector-valued function with differentiable functions f_1, \dots, f_r , and $\Gamma^{(g)}(q_1 \times r)$ is a matrix of unknown parameters. For simplicity, (2) can be expressed as:

$$\eta_i^{(g)} = \Pi^{(g)}G(\eta_i^{(g)}, x_i^{(g)}, \xi_i^{(g)}) + \delta_i^{(g)}$$

where $\delta_i^{(g)}(q_1 \times 1)$ is a vector of error measurements, $\Pi^{(g)} = (B^{(g)}, \Gamma^{(g)})$, and $G(y_i^{(g)}, x_i^{(g)}, \xi_i^{(g)}) = (\eta_i^{(g)T}, F(x_i^{(g)}, \xi_i^{(g)})^T)^T$. It must first be assumed that $\xi_i^{(g)}$ is distributed as $N[0, \Phi]$, further, $\delta_i^{(g)}$ is distributed as $N[0, \Psi_\delta]$, where Ψ_δ is representative of a diagonal matrix containing the elements $\psi_{\delta_1}, \dots, \psi_{\delta_{q_1}}$, and for which $\delta_i^{(g)}$ and $\xi_i^{(g)}$ are each independent.

A more specific example of the generalized nonlinear latent variable defined in (2) that is associated with $\eta_i = (\eta_i)$, $\xi_i = (\xi_{i1}, \xi_{i2})^T$, and $x_i = (x_{i1}, x_{i2})^T$ is:

$$\begin{aligned} \eta_i^{(1)} = & \beta_1^{(1)}x_{i1}^{(1)} + \beta_2^{(1)}x_{i1}^{(1)}\xi_{i1}^{(1)} + \beta_3^{(1)}x_{i1}^{(1)}\xi_{i2}^{(1)} + \beta_4^{(1)}x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)} + \beta_5^{(1)}x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)} + \\ & \gamma_1^{(1)}\xi_{i1}^{(1)} + \gamma_2^{(1)}\xi_{i2}^{(1)} + \\ & \gamma_3^{(1)}\xi_{i1}^{(1)}\xi_{i1}^{(1)} + \gamma_4^{(1)}\xi_{i2}^{(1)}\xi_{i2}^{(1)} + \delta_i^{(1)} \end{aligned} \quad (3)$$

$$\begin{aligned} \eta_i^{(2)} = & \beta_1^{(2)}x_{i1}^{(2)} + \beta_2^{(2)}x_{i1}^{(2)}\xi_{i1}^{(2)} + \beta_3^{(2)}x_{i1}^{(2)}\xi_{i2}^{(2)} + \beta_4^{(2)}x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)} + \beta_5^{(2)}x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)} \\ & + \gamma_1^{(2)}\xi_{i1}^{(2)} + \gamma_2^{(2)}\xi_{i2}^{(2)} + \\ & \gamma_3^{(2)}\xi_{i1}^{(2)}\xi_{i1}^{(2)} + \gamma_4^{(2)}\xi_{i2}^{(2)}\xi_{i2}^{(2)} + \delta_i^{(2)} \end{aligned} \quad (4)$$

Here, $\Gamma^{(1)} = (\beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \beta_4^{(1)}, \beta_5^{(1)}, \gamma_1^{(1)}, \gamma_2^{(1)}, \gamma_3^{(1)}, \gamma_4^{(1)})$,

$$\Gamma^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \beta_4^{(2)}, \beta_5^{(2)}, \gamma_1^{(2)}, \gamma_2^{(2)}, \gamma_3^{(2)}, \gamma_4^{(2)}),$$

and

$$F(x_i^{(g)}, \xi_i^{(g)})$$

$$= (x_{i1}^{(g)}, x_{i1}^{(g)} \xi_{i1}^{(g)}, x_{i1}^{(g)} \xi_{i2}^{(g)}, x_{i1}^{(g)} \xi_{i1}^{(g)} \xi_{i2}^{(g)}, x_{i1}^{(g)} \xi_{i1}^{(g)} \xi_{i2}^{(g)}, \xi_{i1}^{(g)}, \xi_{i2}^{(g)}, \xi_{i1}^{(g)} \xi_{i1}^{(g)}, \xi_{i2}^{(g)} \xi_{i2}^{(g)})^T$$

where $g = 1; 2$. Further, x_i and ξ_i are both quadratic terms of elements, which can be assessed via an appropriately defined LVM. As x_i may be drawn from the arbitrary distributions for covariates that are mixed ordered categorical and dichotomous data, the proposed nonlinear latent variable model can be used to manage a large variety of situations. Furthermore, with regard to nonlinear latent variable models, one must be careful to correctly interpret the mean vector, y_i , namely as it relates to μ .

More specifically, allow A_k and Λ_k to stand for the k^{th} row for each A and Λ .

when $k = 1, \dots, p$; it can be determined, as in Equation (1), that $\mu_{ik} = E(y_{ik})$ which is also $= A_k c_i + \Lambda_k E(w_i)$. However, when $E(\xi_i) = 0$, it can then be determined that in accordance with Equation (2), $E(\eta_i) \neq 0$ if $F(x_i, \xi_i)$ is a nonlinear function of ξ_i . Therefore, going forward, both $E(\omega_i) \neq 0$, and $\mu_{ik} \neq A_k c_i$. So, allow $\Lambda_k = (\Lambda_{k\eta}^T, \Lambda_{k\xi}^T)^T$ to be a partition of Λ_k which corresponds with $\omega_i = (\eta_i^T, \xi_i^T)^T$, which is also a partition. It follows that $E(\xi_i) = 0$, and $\eta_i = (I - B)^{-1} \Gamma F(x_i, \xi_i)$, it follows from Equation (1) that

$$\mu_{ik} = A_k c_i + \Lambda_{k\eta} E(\eta_i) = A_k c_i + \Lambda_{k\eta} [(I - B)^{-1} \Gamma] E(F(x_i, \xi_i)). \quad (5)$$

However, $F(x_i, \xi_i)$ is generally uncomplicated when used in the practical application setting, and as such it can be expected that $E(F(x_i, \xi_i))$ is also relatively simplistic, allowing the computation of μ_{ik} to be performed without any struggle. It is also valuable to study this indirect method for modeling covariates, similar to those demonstrated above, by first supplementing y_i with x_i , and then by then managing each element of the latter as if it is exogenous latent variable which can be measured precisely with use of a single indicator. To manage the difficulties that arise as the result of mixed ordered categorical and dichotomous outcomes, assume that $y_i^{(g)}$ is a $s \times 1$ sub-vector of unobservable continuous responses, and that the information derived from it is reflected by an observable ordered categorical vector $z_i^{(g)}$. Generally speaking, in keeping with this idea, the ordered categorical variable, in this case $z_m^{(g)}$, can be defined according to the related latent continuous random variable $y_m^{(g)}$ by:

$$z^{(g)} = \begin{bmatrix} z_1^{(g)} \\ \vdots \\ z_s^{(g)} \end{bmatrix} \text{ if } \begin{bmatrix} \alpha_{1,z_1^{(g)}}^{(g)} < y_1^{(g)} \leq \alpha_{1,z_1^{(g)}+1}^{(g)} \\ \vdots \\ \alpha_{s,z_s^{(g)}}^{(g)} < y_s^{(g)} \leq \alpha_{s,z_s^{(g)}+1}^{(g)} \end{bmatrix} \quad (6)$$

such that it is also true that $\{-\infty = \alpha_{m,1}^{(g)} < \alpha_{m,2}^{(g)} < \dots < \alpha_{m,b_m}^{(g)} < \alpha_{m,b_m+1}^{(g)} = \infty\}$ is the set of threshold parameters which define the given categories, and b_m for which b_m stands for the number of categories for the ordered categorical variable $z_m^{(g)}$.

Also, when regarding the dichotomous data, the correlation between y and z can be determined according to the set of cut points in which x in Equation (8) signifies dichotomous variables, it follows that

$$z_{ik}^{(g)} = \begin{cases} 1 & \text{if } u_r^{(g)} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

To handle the problem of mixed ordered categorical and dichotomous data in covariates, we will use hidden continuous normal distribution $Q_m^{(g)}$ (truncated normal distribution with known parameters). For ordered categorical covariates data, it follows that:

$$x^{(g)} = \begin{bmatrix} x_1^{(g)} \\ \vdots \\ x_s^{(g)} \end{bmatrix} \text{ if } \begin{bmatrix} \alpha_{1,x_1^{(g)}}^{(g)} < Q_1^{(g)} \leq \alpha_{1,x_1^{(g)}+1}^{(g)} \\ \vdots \\ \alpha_{s,x_s^{(g)}}^{(g)} < Q_s^{(g)} \leq \alpha_{s,x_s^{(g)}+1}^{(g)} \end{bmatrix} \quad (8)$$

To handle the problem of dichotomous data in covariates $X^{(g)}$, we will use hidden continuous normal distribution $Q_m^{(g)}$ (truncated normal distribution with known parameters). It follows that:

$$X_{ik}^{(g)} = \begin{cases} 1 & \text{if } j_r^{(g)} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

It should be understood, however, that for every mixed ordered categorical and dichotomous variable, the number of thresholds (cut points) for each group are equivalent. However, we use two types of thresholds (equal and unequal categories distances).

Lee et. al., (1990) determined that a single-sample method could be implemented with mixed ordered categorical and dichotomous variables, but could not be identified with imposing special identification conditions.

3. Bayesian Analysis of Multi-sample Nonlinear Latent variable Models

Allow $\theta^{(g)}$ be an unknown parameter vector within the previously acknowledged model and similarly allow $\alpha^{(g)}$ be a vector of unknown thresholds for the mixed ordered categorical and dichotomous variables, which correspond to the g th group. This is selected because a select kind of parameter in $\theta^{(g)}$ is frequently identified as an invariant within group models in multi-sample analysis. For example, restrictions on cut points are subject to the following constraints:

$\Lambda^{(1)} = \dots = \Lambda^{(G)}$, $\Phi^{(1)} = \dots = \Phi^{(G)}$ and/or $\Gamma^{(1)} = \dots = \Gamma^{(G)}$, are frequently implemented as thresholds on the model. Hence, when analyzing the data, we can allow certain common parameters, $\theta^{(1)} = \dots = \theta^{(G)}$. More specifically, we allow θ to be a vector containing the complete set of unknown distinct parameters $\Gamma^{(1)} = \dots = \Gamma^{(G)}$, and allow α to be the vector that encompasses all the unknown thresholds.

Thus, the Bayesian estimate of θ and α are generated according to the Gibbs sampler.

More specifically, allow $Z^{(g)} = (z_1^{(g)}, \dots, z_r^{(g)})$ to stand for the observed ordered categorical data matrix, $Z^{(g)} = (z_1^{(g)}, \dots, z_s^{(g)})$ to represent the observed dichotomous data, and $Y^{(g)} = (y_1^{(g)}, \dots, y_n^{(g)})$ and $\Omega^{(g)} = (\omega_1^{(g)}, \dots, \omega_n^{(g)})$ to denote the matrices of latent variables and continuous measurements, respectively.

Then, in the posterior analysis, augment the observed data with Y . Once Y has been established, all the data is accounted for, and considered continuous, so the problem will be simpler to manage. Also, be aware that the observation of Ω , or the nonlinear structural and measurement equations to condense into a regular simultaneous regression model. Complications that arise as the result of the nonlinear relationships between the latent variables are significantly improved.

Thus, issues associated with the more complex elements of the model can be dealt with by augmenting the data. Through posterior analysis, (Z) , which represents the set of observed data, can be supplemented by (Y, Ω) . Further, we will demonstrate the joint posterior distribution $[\theta, \alpha, Y, \Omega | Z]$. The Gibbs sampler, as developed by Geman and

Geman (1984), can be applied in order to create a series of observations from the related joint posterior distribution.

As a result the Bayesian solution can be gained through a series of standard inferences based on the generated sample of observations. Further, by using the Gibbs sampler, we can use the iteration approach to create a set of sample observations from these conditional distributions: $[\Omega|\theta, \alpha, Y, Z]$, $[\alpha, Y|\theta, \Omega, Z]$ and $[\theta|\alpha, Y, \Omega, Z]$. In similar approach to the way previous cut point problems were handled, we can determine the non-informative previous to determining α such that the corresponding prior distribution is comparative to the constant. The conditional distribution $[\theta|\alpha, Y, \Omega, Z]$ thus can be additionally decomposed into individual components which include a number of different structural parameters, according to the various group models.

Some examples of non-nested competing models are:

$$\begin{aligned} M_A: \text{No constraints } M_1 &= \mu^{(1)} = \dots = \mu^{(G)}, M_2 = \Lambda^{(1)} = \dots = \Lambda^{(G)} \\ M_3 &= \Lambda_{\omega}^{(1)} = \dots = \Lambda_{\omega}^{(G)}, M_4 = \Phi_{\omega}^{(1)} = \dots = \Phi_{\omega}^{(G)} \\ M_5 &= \Psi_{\varepsilon}^{(1)} = \dots = \Psi_{\varepsilon}^{(G)}, M_6 = \Psi_{\delta}^{(1)} = \dots = \Psi_{\delta}^{(G)} \end{aligned}$$

These components become varied as different hypotheses are applied, or as competing methodologies are pursued. The components of the conditional distribution, known as $[\theta|\alpha, Y, \Omega, Z]$ and the condition applied to prior distributions are somewhat dissimilar, or varying, when placed under various definitions of M_k as outlined above.

To begin, the prior distributions for the unconstrained parameters from various groups are implicitly expected to act independently. Further, when creating an estimate for the unconstrained parameters, it becomes necessary to identify the specific value of its prior distribution, and to outline the data that belongs in the corresponding groupings, in order for them to be fully applied. In the case of constrained parameters across groups, a prior distribution for the associated constrained parameters is required, and all the related data groups are then combined for estimating (see Song and Lee, 2001).

This section describes the Bayesian estimation and model comparison in the setting of multiple group nonlinear LVMS with ordered categorical variables. In completing the general scheme, the idea of data augmentation is used together with MCMC tools. Theoretically, a multiple group nonlinear LVM is a particular case of the two-level LVM, with some conditional distributions required in the Gibbs sampler which can be achieved from the outcomes. However, as specific constraints among the parameters in different groups are compulsory, it is essential to pay more attention in stipulating the equivalent prior distributions. Likewise, the model comparison in two-level LVMS requires some insight in applying the path sampling procedure (Lee and Song, 2012).

The goal for this section is to define how to analyze the preceding nonlinear LVM, in the context of the mixed ordered categorical and dichotomous variables, using the Bayesian approach. Using this approach is beneficial to the overall application in several ways, including: (1) application of the prior knowledge can enhance the overall analysis when it is directly incorporated. More specifically, it generates more accurate parameter estimation. (2) As demonstrated by various researchers, sampling-based Bayesian methods are not reliant on asymptotic theory (Lee (2006); Lee *et al.* (2010); Lee *et al.*, (2007); Song *et al.*, (2011) (3) Both Bayesian and ML estimates feature similar optimal asymptotic properties. Understanding these principles will allow us to strategically augment the data as termed in the Bayesian estimation of the latent variable models with mixed ordered categorical and dichotomous variables.

When using a Bayesian approach, θ is regarded as the parameter of prior distribution, and a prior density function, and can be defined as $p(\theta)$. The inferences are then based on observed data set, represented by Z_r , Z_s and $p(\theta)$. So let $p(Z, V, \theta)$ stand for the joint probability density function of Z , V and θ under the variety of M_k .

Based on a well-known identity in probability, $p(Z, V, \theta) = p(Z_r, Z_s | \theta)p(\theta)$, where $p(Z_r, Z_s | \theta)$ and $p(\theta | Z_r, Z_s)$ are conditional density functions. It follows that:

$$\log p(\theta | Z_r, Z_s) \propto \log p(Z_r, Z_s | \theta) + \log p(\theta) \quad (10)$$

The function $p(\theta | Z_r, Z_s)$ is called the posterior density function of the unknown parameters.

This yields the posterior density function $p(\theta | Z_r, Z_s)$, or unknown parameters. It also employs the use of sample information and prior knowledge, via the likelihood function $p(Z_r, Z_s | \theta)$, and the prior density function $p(\theta)$. It should be noted, however, that $p(Z_r, Z_s | \theta)$ is dependent on sample size, where $p(\theta)$ is not. As such, for problems with a large sample, $p(\theta)$ is less significant and $p(\theta | Z_r, Z_s)$, the posterior density function, is more relevant, as it is most similar to the likelihood function $p(Z_r, Z_s | \theta)$. 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 contains mixed ordered categorical and dichotomous data. In this case, MCMC methods are applied by allow y_i to be the unobserved variables that parallel the manifest mixed ordered categorical and dichotomous variables in Z_r , Z_s . Also allow that $Y = (y_1, \dots, y_n)$, $U = (u_1, \dots, u_n)$ and $\Omega = (w_1, \dots, w_n)$ are unobserved continuous variables Y, U and latent variables Ω respectively.

Also, one must allow $Y = (y_1, \dots, y_n)$ and $\Omega = (w_1, \dots, w_n)$ when drawing a sufficient and generally large number of observations. If we can draw a sufficiently large number of observations, represented by $\{(\theta^{(t)}, \Omega^{(t)}, Y^{(t)}, U^{(t)}); t = 1, \dots, T\}$ from the joint posterior distribution, defined by $p(\theta, \Omega, Y, U | Z_r, Z_s)$, then the Bayesian estimate for θ as well as any standard error estimates can be derived from the sample mean and variance matrices, respectively.

$$\hat{\theta} = T^{-1} \sum_{t=1}^T \theta^{(t)}, \hat{v} = T^{-1} \sum_{t=1}^T (\theta^{(t)} - \hat{\theta})(\theta^{(t)} - \hat{\theta})'. \quad (11)$$

This means that it is necessary to specifically identify the prior distribution for the related components in θ , even if developing the conditional distribution, $(\theta | \Omega, Y, U, Z_r, Z_s)$ described in Step (1). 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 latent variable models, (see Song and Lee, (2007)). Hence, the following well-known conjugate prior distributions are used:

$$p(A_k) \sim N[A_{0k}, H_{0ak}], p(\lambda_k) \sim N[\lambda_{0k}, H_{0k}], p(\lambda_{\xi k} | \psi_{\delta k}) \sim N[\lambda_{0\xi k}, \psi_{\delta k} H_{0\xi k}], \\ p(\Phi^{-1}) \sim W_q[R_0, \rho_0], p(\psi_{\delta k}^{-1}) \sim \text{Gamma}[\alpha_{0k}, \beta_{0k}] \quad (11)$$

Given the definition that $p(\cdot) \sim$ 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} , λ_k' and $\lambda_{\xi k}'$ are the k^{th} rows of Λ and Λ_{ξ} , respectively. H_{0ak}, H_{0k} and $A_{0k}, \lambda_{0k}, \lambda_{0\xi k}, \alpha_{0k}, \beta_{0k}, \rho_0, \sigma_{0k}, H_{0\xi k}$, 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.

More specifically, when using the Bayesian approach, it is necessary to evaluate the posterior distribution $[\alpha, \theta, \Omega | Z_r, Z_s, Q]$ with the case of non-normal covariates x , 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 mixed ordered categorical and dichotomous variables and covariates in this case, the related conditional distributions can be made too complex to easily derive or simulating data from them. This encourages the additional augmentation of Y, U, X_r, X_s the latent matrices, in the posterior analysis, and motivates attention to the joint posterior distribution $[\alpha, \theta, \Omega, Y, U, X_r, X_s | Z_r, Z_s, Q, J]$. To garner observations of this posterior distribution, using the Gibbs sampler, it is essential to begin with the starting values $(\alpha^{(0)}, \theta^{(0)}, \Omega^{(0)}, Y^{(0)}, U^{(0)}, X_r^{(0)}, X_s^{(0)})$. The following procedure is then implemented to simulate $(\alpha^{(1)}, \theta^{(1)}, \Omega^{(1)}, Y^{(1)}, U^{(1)}, X_r^{(1)}, X_s^{(1)})$ and so on. More specifically at the m^{th} reiteration of the current values $\alpha^{(m)}, \theta^{(m)}, \Omega^{(m)}, Y^{(m)}, U^{(m)}, X_r^{(m)}, X_s^{(m)}$.

1. Generate $\Omega^{(m+1)}$ from $p(\Omega | \theta^{(m+1)}, \alpha^{(m)}, Y^{(m)}, U^{(m)}, X_r^{(m)}, X_s^{(m)}, Z_r^{(m)}, Z_s^{(m)}, Q, J)$
2. Generate $\theta^{(m+1)}$ from $p(\theta | \Omega^{(m+1)}, \alpha^{(m)}, Y^{(m)}, U^{(m)}, X_r^{(m)}, X_s^{(m)}, Z_r^{(m)}, Z_s^{(m)}, Q, J)$
3. Generate $(\alpha^{(m+1)}, Y^{(m+1)}, X_r^{(m+1)}, X_s^{(m+1)})$ from $p(\alpha, Y, U, X_r, X_s | \theta^{(m+1)}, \Omega^{(m+1)}, Z_r, Z_s, Q, J)$

(12)

The cycle, as previously defined, will give us $(\alpha^{(m+1)}, \theta^{(m+1)}, \Omega^{(m+1)}, Y^{(m+1)}, U^{(m+1)}, X_r^{(m)}, X_s^{(m)})$, only occurring after the m th repetition. So, as m approaches infinity, the joint distribution of the value of $(\alpha^{(m)}, \theta^{(m)}, \Omega^{(m)}, Y^{(m)}, U^{(m)}, X_r^{(m)}, X_s^{(m)})$ can be proven to move toward the joint posterior distribution $[\alpha, \theta, \Omega, Y, X_r, X_s | Z, Q, J]$. (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. 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:

$$DIC_k = \overline{D(\theta_k)} + d_k, \quad (14)$$

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

$$\overline{D(\theta_k)} = E_{\theta_k} \{-2 \log p(Z | \theta_k, M_k) | Z\}. \quad (15)$$

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

$$d_k = E_{\theta_k} [-2 \log p(Z | \theta_k, M_k) | Z] + 2 \log f(Z | \tilde{\theta}). \quad (16)$$

in which $\tilde{\theta}$ is the Bayesian estimate of θ . Let $\{\theta_k^{(t)}: t = 1, \dots, T\}$ be a sample of observations simulated from the posterior distribution. The expectations in Equations (15) and (16) can be estimated as follows:

$$E_{\theta_k}\{-2 \log p(Z|\theta_k, M_k)|Z\} = -\frac{2}{T} \sum_{t=1}^T \log p(Z|\theta_k^{(t)}, M_k). \quad (17)$$

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

5. A Case Study of Psychological data

Let us consider the data which can be used for deriving results of two different independent samples which are selected from the natural history based on the study of a rural drug found in Ohio (n=200) and Kentucky (n=200) in USA between the time 2003 and 2005 (Booth, et. al, 2006).

Further many amendments were made under the BSI-18 scale which involved 3 categories of psychiatric disorders which was measured under this scale and considered aspects like somatization (SOM), depression (DEP), and anxiety (ANX).

The data consists of sixteen variables and two covariates in each group. Moreover 9 of these variables were evaluated based on ordered categorical variables: (1, not at all; 2, a little bit; 3, moderately; 4, quite a bit; 5, extremely and the rest of variables are dichotomous variables and recoded as a dummy variable: 0, 'not at all' or 'a little bit,' 1, 'Moderately' to 'Extremely' for the purpose of model demonstration (Wang and Wang, 2012).

A real data study is presented here to give some idea of the empirical performance of the proposed Bayesian approach in which 16 manifest variables are related to two basic latent variables $(\eta^{(g)}, \xi_{i1}^{(g)} \xi_{i2}^{(g)})$ from multi-sample nonlinear LVMs defined in Equation (19) and Equation (20), respectively. Hence, some quadratic and interaction effects of the latent variables are considered. To illustrate the Bayesian methods in analysing linear and nonlinear latent variable models with mixed ordered categorical and dichotomous variables, we use a real data set that is related to random vectors with $G=1,2$, $z_i^{(g)} = (z_{i1}^{(g)}, z_{i2}^{(g)}, \dots, z_{i16}^{(g)})'$, let $y_i^{(g)} = (y_{i1}^{(g)}, y_{i2}^{(g)}, \dots, y_{i16}^{(g)})'$ be the latent continuous random vector, which corresponds to the mixed ordered categorical and dichotomous variables $z_{i1}^{(g)}, z_{i2}^{(g)}, \dots, z_{i16}^{(g)}$ where $z_i^{(g)}, i = 1, \dots, n$ are mixed ordered categorical and dichotomous variables that are related to (3) latent variables $w_i^{(g)} = (\eta_i^{(g)}, \xi_{i1}^{(g)}, \xi_{i2}^{(g)})$, $\varepsilon_i^{(g)} = (\varepsilon_{i1}^{(g)}, \varepsilon_{i2}^{(g)}, \dots, \varepsilon_{i16}^{(g)})$, with the following values of the parameters in $\Lambda^{(g)} = (\lambda_1^{(g)}, \lambda_2^{(g)}, \dots, \lambda_{13}^{(g)})'$

$$\begin{aligned} &= \Lambda^{(1)} \begin{bmatrix} 1^* & \lambda_{21}^{(1)} & \lambda_{31}^{(1)} \lambda_{41}^{(1)} & \lambda_{51}^{(1)} & \lambda_{61}^{(1)} 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1 & \lambda_{82}^{(1)} & \lambda_{92}^{(1)} \lambda_{102}^{(1)} & \lambda_{112}^{(1)} & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1 & \lambda_{133}^{(1)} & \lambda_{143}^{(1)} \lambda_{153}^{(1)} & \lambda_{163}^{(1)} \end{bmatrix}, \\ &\Lambda^{(2)} \begin{bmatrix} 1^* & \lambda_{21}^{(2)} & \lambda_{31}^{(2)} \lambda_{41}^{(2)} & \lambda_{51}^{(2)} & \lambda_{61}^{(2)} 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1 & \lambda_{82}^{(2)} & \lambda_{92}^{(2)} \lambda_{102}^{(2)} & \lambda_{112}^{(2)} & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* & 1 & \lambda_{133}^{(2)} & \lambda_{143}^{(2)} \lambda_{153}^{(2)} & \lambda_{163}^{(2)} \end{bmatrix}, \\ &\Phi^{(g)} = \Phi^{(1)} \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{bmatrix}, \Phi^{(2)} \begin{bmatrix} \phi_{11}^{(2)} & \phi_{12}^{(2)} \\ \phi_{21}^{(2)} & \phi_{22}^{(2)} \end{bmatrix}. \end{aligned} \quad (18)$$

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

The relationships of the latent variables in $w_i^{(g)} = (\eta_i^{(g)}, \xi_{i1}^{(g)} \xi_{i2}^{(g)})$ are assessed by the nonlinear latent variable, which is described in the following equations.

$$\eta_i^{(1)} = \beta_1^{(1)} x_{i1}^{(1)} + \beta_2^{(1)} x_{i2}^{(1)} + \beta_3^{(1)} x_{i1}^{(1)} x_{i2}^{(1)} + \beta_4^{(1)} x_{i1}^{(1)} \xi_{i2}^{(1)} + \beta_5^{(1)} x_{i1}^{(1)} \xi_{i1}^{(1)} \xi_{i2}^{(1)} + \gamma_1^{(1)} \xi_{i1}^{(1)} + \gamma_2^{(1)} \xi_{i2}^{(1)} + \gamma_3^{(1)} \xi_{i1}^{(1)} \xi_{i1}^{(1)} + \gamma_4^{(1)} \xi_{i2}^{(1)} \xi_{i2}^{(1)} + \delta_i^{(1)} \quad (19)$$

$$\eta_i^{(2)} = \beta_1^{(2)} x_{i1}^{(2)} + \beta_2^{(2)} x_{i2}^{(2)} + \beta_3^{(2)} x_{i1}^{(2)} x_{i2}^{(2)} + \beta_4^{(2)} x_{i1}^{(2)} \xi_{i2}^{(2)} + \beta_5^{(2)} x_{i1}^{(2)} \xi_{i1}^{(2)} \xi_{i2}^{(2)} + \gamma_1^{(2)} \xi_{i1}^{(2)} + \gamma_2^{(2)} \xi_{i2}^{(2)} + \gamma_3^{(2)} \xi_{i1}^{(2)} \xi_{i1}^{(2)} + \gamma_4^{(2)} \xi_{i2}^{(2)} \xi_{i2}^{(2)} + \delta_i^{(2)} \quad (20)$$

Here, $\Gamma^{(1)} = (\beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \beta_4^{(1)}, \beta_5^{(1)}, \gamma_1^{(1)}, \gamma_2^{(1)}, \gamma_3^{(1)}, \gamma_4^{(1)})$,

and $\Gamma^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \beta_4^{(2)}, \beta_5^{(2)}, \gamma_1^{(2)}, \gamma_2^{(2)}, \gamma_3^{(2)}, \gamma_4^{(2)})$,

and

$F(x_i^{(g)}, \xi_i^{(g)}) = (x_{i1}^{(g)}, x_{i2}^{(g)}, x_{i1}^{(g)} x_{i2}^{(g)}, x_{i1}^{(g)} \xi_{i2}^{(g)}, x_{i1}^{(g)} \xi_{i1}^{(g)} \xi_{i2}^{(g)}, \xi_{i1}^{(g)}, \xi_{i2}^{(g)}, \xi_{i1}^{(g)} \xi_{i1}^{(g)}, \xi_{i2}^{(g)} \xi_{i2}^{(g)})^T$. The covariates X_i come from arbitrary distributions that give mixed ordered categorical and dichotomous data.

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 (12) are set equal to the following values with initial values equal to 0 in some parameters and 1 in other parameters for two groups of data;

$R_0^{-1} = 8\Phi$, 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 Equation (12) are set equal to the following values with initial values equal to 0.5 for two groups of data;

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

6. Results and Discussion

The objective of this section is to present results of a simulation study to reveal the empirical performance of the Bayesian estimates and the DIC for model comparison.

For nonlinear LVMs with covariates, we have the following proposed models for $g=1,2$:

$$\begin{aligned} M_1: \eta_i^{(g)} &= \beta_1^{(g)} x_{i1}^{(g)} + \beta_2^{(g)} x_{i1}^{(g)} \xi_{i1}^{(g)} + \beta_3^{(g)} x_{i1}^{(g)} \xi_{i2}^{(g)} + \gamma_1^{(g)} \xi_{i1}^{(g)} + \gamma_2^{(g)} \xi_{i2}^{(g)} + \delta_i^{(g)} \\ M_2: \eta_i^{(g)} &= \beta_1^{(g)} x_{i1}^{(g)} + \beta_2^{(g)} x_{i1}^{(g)} \xi_{i1}^{(g)} + \beta_3^{(g)} x_{i1}^{(g)} \xi_{i2}^{(g)} + \beta_4^{(g)} x_{i1}^{(g)} x_{i1}^{(g)} \xi_{i2}^{(g)} + \gamma_1^{(g)} \xi_{i1}^{(g)} + \gamma_2^{(g)} \xi_{i2}^{(g)} + \delta_i^{(g)} \\ M_3: \eta_i^{(g)} &= \beta_1^{(g)} x_{i1}^{(g)} + \beta_2^{(g)} x_{i1}^{(g)} \xi_{i2}^{(g)} + \beta_3^{(g)} x_{i1}^{(g)} \xi_{i1}^{(g)} \xi_{i1}^{(g)} + \gamma_1^{(g)} \xi_{i1}^{(g)} + \gamma_2^{(g)} \xi_{i2}^{(g)} + \gamma_3^{(g)} \xi_{i1}^{(g)} \xi_{i1}^{(g)} + \gamma_4^{(g)} \xi_{i2}^{(g)} \xi_{i2}^{(g)} + \delta_i^{(g)} \\ M_4: \eta_i^{(1)} &= \beta_1^{(1)} x_{i1}^{(1)} + \beta_2^{(1)} x_{i2}^{(1)} + \beta_3^{(1)} x_{i1}^{(1)} x_{i2}^{(1)} + \beta_4^{(1)} x_{i1}^{(1)} \xi_{i2}^{(1)} + \beta_5^{(1)} x_{i1}^{(1)} \xi_{i1}^{(1)} \xi_{i2}^{(1)} + \gamma_1^{(1)} \xi_{i1}^{(1)} + \gamma_2^{(1)} \xi_{i2}^{(1)} + \gamma_3^{(1)} \xi_{i1}^{(1)} \xi_{i1}^{(1)} + \gamma_4^{(1)} \xi_{i2}^{(1)} \xi_{i2}^{(1)} + \delta_i^{(1)} \end{aligned}$$

TABLE 1. Bayesian Estimation of Nonlinear LVMs with Mixed Variables under Prior I and II Using Censored Normal Distribution with Equal Spaces of Thresholds

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\lambda_{21}^{(1)}$	0.907	0.112	[0.699,1.143]	$\phi_{11}^{(1)}$	0.386	0.068	[0.268,0.537]
$\lambda_{31}^{(1)}$	0.688	0.096	[0.511,0.886]	$\phi_{12}^{(1)}$	0.297	0.053	[0.202,0.412]
$\lambda_{41}^{(1)}$	0.833	0.135	[0.583,1.114]	$\phi_{22}^{(1)}$	0.297	0.053	[0.202,0.412]
$\lambda_{51}^{(1)}$	0.874	0.115	[0.663,1.117]	$\phi_{11}^{(2)}$	0.397	0.068	[0.282,0.549]
$\lambda_{61}^{(1)}$	0.804	0.100	[0.617,1.007]	$\phi_{12}^{(2)}$	0.459	0.088	[0.318,0.671]
$\lambda_{82}^{(1)}$	0.939	0.156	[0.652,1.262]	$\phi_{22}^{(2)}$	0.360	0.067	[0.248,0.507]
$\lambda_{92}^{(1)}$	1.114	0.216	[0.720,1.553]	$\gamma_1^{(1)}$	0.680	0.256	[0.177, 1.180]
$\lambda_{102}^{(1)}$	1.272	0.232	[0.861,1.768]	$\gamma_2^{(1)}$	0.720	0.250	[0.242, 1.211]
$\lambda_{112}^{(1)}$	0.657	0.166	[0.367,1.014]	$\gamma_3^{(1)}$	0.079	0.277	[-0.420, 0.675]
$\lambda_{122}^{(1)}$	0.814	0.175	[0.507,1.192]	$\gamma_4^{(1)}$	0.052	0.257	[-0.417, 0.606]
$\lambda_{143}^{(1)}$	0.815	0.187	[0.497,1.227]	$\gamma_1^{(2)}$	0.791	0.254	[0.313, 1.304]
$\lambda_{153}^{(1)}$	0.397	0.157	[0.144,0.752]	$\gamma_2^{(2)}$	0.609	0.258	[0.119, 1.136]
$\lambda_{163}^{(1)}$	0.767	0.182	[0.453,1.144]	$\gamma_3^{(2)}$	0.187	0.268	[-0.324, 0.708]
$\lambda_{21}^{(2)}$	0.921	0.111	[0.715,1.151]	$\gamma_4^{(2)}$	0.350	0.284	[-0.213, 0.911]
$\lambda_{31}^{(2)}$	0.897	0.114	[0.685,1.139]	$\beta_1^{(1)}$	0.503	0.182	[0.150, 0.854]
$\lambda_{41}^{(2)}$	0.891	0.143	[0.637,1.198]	$\beta_2^{(1)}$	0.439	0.146	[0.168, 0.748]
$\lambda_{51}^{(2)}$	0.905	0.119	[0.686,1.152]	$\beta_3^{(1)}$	0.150	0.199	[-0.237, 0.531]
$\lambda_{61}^{(2)}$	0.896	0.115	[0.686,1.136]	$\beta_4^{(1)}$	-0.246	0.241	[-0.716, 0.234]
$\lambda_{82}^{(2)}$	1.168	0.167	[0.863,1.508]	$\beta_5^{(1)}$	0.230	0.281	[-0.302, 0.808]
$\lambda_{92}^{(2)}$	0.184	0.184	[-0.168,0.555]	$\beta_1^{(2)}$	0.683	0.221	[0.280, 1.135]
$\lambda_{102}^{(2)}$	0.101	0.165	[-0.216,0.435]	$\beta_2^{(2)}$	0.387	0.186	[0.047, 0.767]
$\lambda_{112}^{(2)}$	0.564	0.156	[0.325,0.919]	$\beta_3^{(2)}$	0.076	0.241	[-0.392, 0.549]
$\lambda_{122}^{(2)}$	0.713	0.183	[0.416,1.130]	$\beta_4^{(2)}$	-0.280	0.266	[-0.787, 0.271]
$\lambda_{143}^{(2)}$	0.751	0.177	[0.447,1.127]	$\beta_5^{(2)}$	0.564	0.308	[-0.015, 1.201]
$\lambda_{153}^{(2)}$	0.287	0.114	[0.103,0.545]	$\psi_{\epsilon\delta}^{(1)}$	0.502	0.107	[0.332,0.747]
$\lambda_{163}^{(2)}$	0.667	0.164	[0.385,1.007]	$\psi_{\epsilon\delta}^{(2)}$	0.518	0.121	[0.328,0.800]

TABLE 2. Bayesian Estimation of Nonlinear LVMs with Mixed Variables under Prior I and II Using Censored Normal Distribution with Unequal Spaces of Thresholds

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\lambda_{21}^{(1)}$	0.852	0.116	[0.637,1.095]	$\phi_{11}^{(1)}$	0.299	0.051	[0.210,0.408]
$\lambda_{31}^{(1)}$	0.631	0.095	[0.456,0.830]	$\phi_{12}^{(1)}$	0.227	0.040	[0.157,0.311]
$\lambda_{41}^{(1)}$	0.794	0.143	[0.535,1.097]	$\phi_{22}^{(1)}$	0.227	0.040	[0.157,0.311]
$\lambda_{51}^{(1)}$	0.852	0.121	[0.633,1.104]	$\phi_{11}^{(2)}$	0.335	0.055	[0.243,0.456]
$\lambda_{61}^{(1)}$	0.699	0.098	[0.517,0.894]	$\phi_{12}^{(2)}$	0.365	0.068	[0.250,0.512]
$\lambda_{82}^{(1)}$	0.824	0.147	[0.552,1.127]	$\phi_{22}^{(2)}$	0.269	0.053	[0.179,0.381]
$\lambda_{92}^{(1)}$	1.191	0.225	[0.782,1.658]	$\gamma_1^{(1)}$	0.710	0.259	[0.216,1.236]
$\lambda_{102}^{(1)}$	1.329	0.233	[0.907,1.821]	$\gamma_2^{(1)}$	0.863	0.260	[0.370,1.386]
$\lambda_{112}^{(1)}$	0.676	0.158	[0.407,1.035]	$\gamma_3^{(1)}$	0.065	0.303	[-0.507,0.690]
$\lambda_{122}^{(1)}$	0.824	0.184	[0.523,1.224]	$\gamma_4^{(1)}$	0.036	0.266	[-0.474,0.577]
$\lambda_{143}^{(1)}$	0.875	0.202	[0.524,1.300]	$\gamma_1^{(2)}$	0.908	0.261	[0.415,1.459]
$\lambda_{153}^{(1)}$	0.404	0.154	[0.148,0.741]	$\gamma_2^{(2)}$	0.683	0.275	[0.165,1.244]
$\lambda_{163}^{(1)}$	0.777	0.180	[0.474,1.179]	$\gamma_3^{(2)}$	0.272	0.309	[-0.399,0.839]
$\lambda_{21}^{(2)}$	0.848	0.113	[0.650,1.097]	$\gamma_4^{(2)}$	0.373	0.273	[-0.153,0.914]
$\lambda_{31}^{(2)}$	0.825	0.119	[0.604,1.076]	$\beta_1^{(1)}$	0.505	0.196	[0.129, 0.911]
$\lambda_{41}^{(2)}$	0.789	0.140	[0.530,1.080]	$\beta_2^{(1)}$	0.456	0.151	[0.166, 0.769]
$\lambda_{51}^{(2)}$	0.864	0.120	[0.647,1.115]	$\beta_3^{(1)}$	0.189	0.209	[-0.221, 0.599]
$\lambda_{61}^{(2)}$	0.826	0.116	[0.612,1.065]	$\beta_4^{(1)}$	-0.251	0.250	[-0.745, 0.238]
$\lambda_{82}^{(2)}$	1.167	0.174	[0.856,1.534]	$\beta_5^{(1)}$	0.299	0.302	[-0.255, 0.956]
$\lambda_{92}^{(2)}$	0.240	0.205	[-0.144,0.669]	$\beta_1^{(2)}$	0.766	0.216	[0.370, 1.222]
$\lambda_{102}^{(2)}$	0.125	0.181	[-0.214,0.506]	$\beta_2^{(2)}$	0.411	0.182	[0.059, 0.781]
$\lambda_{112}^{(2)}$	0.558	0.149	[0.317,0.887]	$\beta_3^{(2)}$	0.064	0.248	[-0.436, 0.533]
$\lambda_{122}^{(2)}$	0.744	0.184	[0.440,1.168]	$\beta_4^{(2)}$	-0.176	0.275	[-0.697, 0.372]
$\lambda_{143}^{(2)}$	0.773	0.193	[0.446,1.165]	$\beta_5^{(2)}$	0.568	0.327	[-0.053,1.247]
$\lambda_{153}^{(2)}$	0.302	0.120	[0.105,0.584]	$\psi_{\epsilon\delta}^{(1)}$	0.497	0.106	[0.329,0.739]
$\lambda_{163}^{(2)}$	0.621	0.160	[0.360,0.984]	$\psi_{\epsilon\delta}^{(2)}$	0.512	0.121	[0.325,0.792]

TABLE 3. Bayesian Estimation of Nonlinear LVMs with Mixed Variables under Prior I and II Using Truncated Normal Distribution with Equal Spaces of Thresholds

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\lambda_{21}^{(1)}$	0.911	0.147	[0.642,1.218]	$\phi_{11}^{(1)}$	0.500	0.102	[0.332,0.729]
$\lambda_{31}^{(1)}$	0.694	0.129	[0.454,0.960]	$\phi_{12}^{(1)}$	0.393	0.075	[0.261,0.553]
$\lambda_{41}^{(1)}$	0.925	0.167	[0.613,1.264]	$\phi_{22}^{(1)}$	0.393	0.075	[0.261,0.553]
$\lambda_{51}^{(1)}$	0.845	0.146	[0.575,1.145]	$\phi_{11}^{(2)}$	0.520	0.093	[0.355,0.719]
$\lambda_{61}^{(1)}$	0.807	0.133	[0.561,1.078]	$\phi_{12}^{(2)}$	0.572	0.102	[0.394,0.798]
$\lambda_{82}^{(1)}$	0.917	0.164	[0.614,1.252]	$\phi_{22}^{(2)}$	0.455	0.089	[0.299,0.648]
$\lambda_{92}^{(1)}$	1.193	0.218	[0.803,1.662]	$\gamma_1^{(1)}$	0.505	0.232	[0.046,0.966]
$\lambda_{102}^{(1)}$	1.280	0.221	[0.884,1.751]	$\gamma_2^{(1)}$	0.612	0.233	[0.145,1.068]
$\lambda_{112}^{(1)}$	0.737	0.147	[0.484,1.053]	$\gamma_3^{(1)}$	0.001	0.236	[-0.450,0.446]
$\lambda_{122}^{(1)}$	0.884	0.173	[0.590,1.260]	$\gamma_4^{(1)}$	-0.011	0.252	[-0.502,0.491]
$\lambda_{143}^{(1)}$	0.935	0.182	[0.614,1.329]	$\gamma_1^{(2)}$	0.666	0.247	[0.195,1.157]
$\lambda_{153}^{(1)}$	0.489	0.178	[0.174,0.880]	$\gamma_2^{(2)}$	0.506	0.249	[0.032,1.007]
$\lambda_{163}^{(1)}$	0.882	0.172	[0.584,1.262]	$\gamma_3^{(2)}$	0.131	0.270	[-0.389,0.644]
$\lambda_{21}^{(2)}$	0.970	0.143	[0.707,1.271]	$\gamma_4^{(2)}$	0.193	0.249	[-0.292,0.714]
$\lambda_{31}^{(2)}$	0.935	0.140	[0.676,1.221]	$\beta_1^{(1)}$	0.490	0.188	[0.132, 0.869]
$\lambda_{41}^{(2)}$	0.947	0.156	[0.657,1.271]	$\beta_2^{(1)}$	0.395	0.131	[0.153, 0.682]
$\lambda_{51}^{(2)}$	0.921	0.137	[0.674,1.203]	$\beta_3^{(1)}$	0.166	0.183	[-0.187, 0.514]
$\lambda_{61}^{(2)}$	0.895	0.143	[0.627,1.198]	$\beta_4^{(1)}$	-0.211	0.223	[-0.647, 0.229]
$\lambda_{82}^{(2)}$	1.225	0.189	[0.871,1.611]	$\beta_5^{(1)}$	0.187	0.261	[-0.284, 0.741]
$\lambda_{92}^{(2)}$	0.204	0.172	[-0.119,0.554]	$\beta_1^{(2)}$	0.693	0.197	[0.321, 1.096]
$\lambda_{102}^{(2)}$	0.111	0.155	[-0.187,0.422]	$\beta_2^{(2)}$	0.386	0.173	[0.056, 0.735]
$\lambda_{112}^{(2)}$	0.583	0.120	[0.368,0.824]	$\beta_3^{(2)}$	0.090	0.232	[-0.354, 0.554]
$\lambda_{122}^{(2)}$	0.793	0.161	[0.510,1.142]	$\beta_4^{(2)}$	-0.292	0.242	[-0.763, 0.176]
$\lambda_{143}^{(2)}$	0.826	0.149	[0.555,1.137]	$\beta_5^{(2)}$	0.517	0.268	[0.014,1.068]
$\lambda_{153}^{(2)}$	0.357	0.120	[0.151,0.631]	$\psi_{\epsilon\delta}^{(1)}$	0.517	0.268	[0.014, 1.068]
$\lambda_{163}^{(2)}$	0.705	0.158	[0.428,1.042]	$\psi_{\epsilon\delta}^{(2)}$	0.441	0.091	[0.294, 0.650]

TABLE 4. Bayesian Estimation of Nonlinear LVMs with Mixed Variables under Prior I and II Using Truncated Normal Distribution with Unequal Spaces of Thresholds

Para	Est.	SE	HPD Interval	Para	Est.	SE	HPD Interval
$\lambda_{21}^{(1)}$	0.852	0.155	[0.565,1.170]	$\phi_{11}^{(1)}$	0.432	0.095	[0.277,0.648]
$\lambda_{31}^{(1)}$	0.655	0.139	[0.403,0.946]	$\phi_{12}^{(1)}$	0.322	0.068	[0.203,0.468]
$\lambda_{41}^{(1)}$	0.894	0.174	[0.569,1.252]	$\phi_{22}^{(1)}$	0.322	0.068	[0.203,0.468]
$\lambda_{51}^{(1)}$	0.802	0.151	[0.523,1.113]	$\phi_{11}^{(2)}$	0.445	0.086	[0.300,0.642]
$\lambda_{61}^{(1)}$	0.762	0.141	[0.503,1.054]	$\phi_{12}^{(2)}$	0.507	0.108	[0.334,0.751]
$\lambda_{82}^{(1)}$	0.896	0.178	[0.570,1.268]	$\phi_{22}^{(2)}$	0.370	0.086	[0.233,0.568]
$\lambda_{92}^{(1)}$	1.278	0.234	[0.840,1.760]	$\gamma_1^{(1)}$	0.570	0.247	[0.114,1.079]
$\lambda_{102}^{(1)}$	1.322	0.232	[0.885,1.791]	$\gamma_2^{(1)}$	0.662	0.240	[0.190,1.122]
$\lambda_{112}^{(1)}$	0.756	0.151	[0.488,1.070]	$\gamma_3^{(1)}$	-0.029	0.250	[-0.495,0.477]
$\lambda_{122}^{(1)}$	0.891	0.168	[0.599,1.250]	$\gamma_4^{(1)}$	0.073	0.260	[-0.430,0.595]
$\lambda_{143}^{(1)}$	0.940	0.187	[0.607,1.338]	$\gamma_1^{(2)}$	0.744	0.243	[0.284,1.228]
$\lambda_{153}^{(1)}$	0.466	0.163	[0.175,0.817]	$\gamma_2^{(2)}$	0.517	0.253	[0.025,1.014]
$\lambda_{163}^{(1)}$	0.907	0.185	[0.585,1.304]	$\gamma_3^{(2)}$	0.205	0.257	[-0.285,0.746]
$\lambda_{21}^{(2)}$	0.911	0.148	[0.639,1.219]	$\gamma_4^{(2)}$	0.291	0.247	[-0.192,0.779]
$\lambda_{31}^{(2)}$	0.872	0.147	[0.601,1.180]	$\beta_1^{(1)}$	0.453	0.194	[0.081,0.855]
$\lambda_{41}^{(2)}$	0.904	0.171	[0.592,1.258]	$\beta_2^{(1)}$	0.428	0.140	[0.168,0.719]
$\lambda_{51}^{(2)}$	0.875	0.151	[0.589,1.181]	$\beta_3^{(1)}$	0.168	0.202	[-0.224,0.571]
$\lambda_{61}^{(2)}$	0.852	0.151	[0.577,1.170]	$\beta_4^{(1)}$	-0.180	0.250	[-0.678,0.309]
$\lambda_{82}^{(2)}$	1.210	0.195	[0.849,1.610]	$\beta_5^{(1)}$	0.299	0.299	[-0.232,0.911]
$\lambda_{92}^{(2)}$	0.250	0.189	[-0.106,0.646]	$\beta_1^{(2)}$	0.761	0.195	[0.398,1.163]
$\lambda_{102}^{(2)}$	0.141	0.166	[-0.176,0.473]	$\beta_2^{(2)}$	0.376	0.182	[0.027,0.759]
$\lambda_{112}^{(2)}$	0.604	0.130	[0.374,0.878]	$\beta_3^{(2)}$	0.073	0.244	[-0.406,0.569]
$\lambda_{122}^{(2)}$	0.801	0.156	[0.525,1.136]	$\beta_4^{(2)}$	-0.103	0.265	[-0.629,0.408]
$\lambda_{143}^{(2)}$	0.874	0.164	[0.589,1.223]	$\beta_5^{(2)}$	0.448	0.286	[-0.073,1.037]
$\lambda_{153}^{(2)}$	0.349	0.123	[0.142,0.629]	$\psi_{\epsilon\delta}^{(1)}$	0.436	0.089	[0.294,0.638]
$\lambda_{163}^{(2)}$	0.688	0.155	[0.411,1.026]	$\psi_{\epsilon\delta}^{(2)}$	0.489	0.104	[0.322,0.726]

Table 5. Goodness of Fit Statistics (DIC) for LVMs with Mixed Variables and Equal Spaces of Thresholds under Prior I and II

	Interval Censored Normal	Interval Truncated Normal
Group 1, n ₁ =200	6583	7006
Group 1, n ₂ =200	6088	6450
Total	12671	13456

Table 6. Goodness of Fit Statistics (DIC) for LVMs with Mixed Variables and Unequal Spaces of Thresholds under Prior I and II

	Interval Censored Normal	Interval Truncated Normal
Group 1, n ₁ =200	6570	7146
Group 1, n ₂ =200	5838	6321
Total	12408	13467

The estimated multiple group nonlinear latent variable using censored normal distribution with equal spaces of thresholds is given by

$$\eta_i^{(1)} = (0.503)x_{i1}^{(1)} + (0.439)x_{i2}^{(1)} + (0.150)x_{i1}^{(1)}x_{i2}^{(1)} + (-0.246)x_{i1}^{(1)}\xi_{i2}^{(1)} + (0.230)x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)} + (0.680)\xi_{i1}^{(1)} + (0.720)\xi_{i2}^{(1)} + (0.079)\xi_{i1}^{(1)}\xi_{i1}^{(1)} + (0.052)\xi_{i2}^{(1)}\xi_{i2}^{(1)} + \delta_i^{(1)} \quad (21)$$

$$\eta_i^{(2)} = (0.683)x_{i1}^{(2)} + (0.387)x_{i2}^{(2)} + (0.076)x_{i1}^{(2)}x_{i2}^{(2)} + (-0.280)x_{i1}^{(2)}\xi_{i2}^{(2)} + (0.564)x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)} + (0.791)\xi_{i1}^{(2)} + (0.609)\xi_{i2}^{(2)} + (0.187)\xi_{i1}^{(2)}\xi_{i1}^{(2)} + (0.350)\xi_{i2}^{(2)}\xi_{i2}^{(2)} + \delta_i^{(2)} \quad (22)$$

From the standard error estimates and the HPD intervals, we observe that all the linear effects of the covariates and the exogenous latent variables, and the interaction effects corresponding to $x_{i1}^{(1)}x_{i2}^{(1)}, x_{i1}^{(1)}\xi_{i2}^{(1)}, x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)}$ in the first group (Ohio) are quite different from zero, $x_{i1}^{(2)}\xi_{i2}^{(2)}, x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)}, \xi_{i1}^{(2)}\xi_{i1}^{(2)}, \xi_{i2}^{(2)}\xi_{i2}^{(2)}$ in the second group (Kentucky) are quite different from zero, while the quadratic interaction effects corresponding to $\xi_{i1}^{(1)}\xi_{i1}^{(1)}, \xi_{i2}^{(1)}\xi_{i2}^{(1)}$ in first group (Ohio) and $x_{i1}^{(2)}x_{i2}^{(2)}$ in the second group (Kentucky) is small. Hence, the quadratic interaction causal effects of depression ($\xi_{i1}^{(1)}\xi_{i1}^{(1)}$), and also anxiety ($\xi_{i2}^{(1)}\xi_{i2}^{(1)}$) in the first group (Ohio) and $x_{i1}^{(2)}x_{i2}^{(2)}$ in the second group (Kentucky) is not significant when given the other linear and interaction effects in the latent variable. As $\xi_{i1}^{(1)}, \xi_{i2}^{(1)}, x_{i1}^{(1)}, x_{i2}^{(1)}$ and $\xi_{i1}^{(2)}, \xi_{i2}^{(2)}, x_{i1}^{(2)}, x_{i2}^{(2)}$ in the first and second groups is related to the mixed ordered categorical and dichotomous variables, the standard error estimate (0.241) and (0.281) correspond to $\beta_4^{(1)} = (-0.246), \beta_5^{(1)} = (0.230)$ is quite large, and the HPD interval, (-0.716, 0.234) and (-0.302, 0.808) is quite wide. However, as $\beta_4^{(1)} = (-0.246), \beta_5^{(1)} = (0.230)$ is quite different from zero in magnitude. The standard error estimate (0.266), (0.308), (0.268) and (0.284) correspond to $\beta_4^{(2)} = (-0.280), \beta_5^{(2)} = (0.564)$ and $\gamma_3^{(2)} = (0.187), \gamma_4^{(2)} = (0.350)$ is quite large, and the HPD interval, (-0.787, 0.271), (-0.015, 1.201), (-0.324, 0.708) and (-0.213, 0.911) is quite wide. Hence the corresponding interaction effect is substantial.

The estimated multiple group nonlinear latent variable using censored normal distribution with unequal spaces of thresholds is given by

$$\eta_i^{(1)} = (0.505)x_{i1}^{(1)} + (0.456)x_{i2}^{(1)} + (0.189)x_{i1}^{(1)}x_{i2}^{(1)} + (-0.251)x_{i1}^{(1)}\xi_{i2}^{(1)} + (0.299)x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)} + (0.710)\xi_{i1}^{(1)} + (0.863)\xi_{i2}^{(1)} + (0.065)\xi_{i1}^{(1)}\xi_{i1}^{(1)} + (0.036)\xi_{i2}^{(1)}\xi_{i2}^{(1)} + \delta_i^{(1)} \quad (23)$$

$$\eta_i^{(2)} = (0.766)x_{i1}^{(2)} + (0.411)x_{i2}^{(2)} + (0.064)x_{i1}^{(2)}x_{i2}^{(2)} + (-0.176)x_{i1}^{(2)}\xi_{i2}^{(2)} + (0.568)x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)} + (0.908)\xi_{i1}^{(2)} + (0.683)\xi_{i2}^{(2)} + (0.272)\xi_{i1}^{(2)}\xi_{i1}^{(2)} + (0.373)\xi_{i2}^{(2)}\xi_{i2}^{(2)} + \delta_i^{(2)} \quad (24)$$

From the standard error estimates and the HPD intervals, we observe that all the linear effects of the covariates and the exogenous latent variables, and the interaction effects corresponding to $x_{i1}^{(1)}x_{i2}^{(1)}, x_{i1}^{(1)}\xi_{i2}^{(1)}, x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)}$ in the first group (Ohio) are quite different from zero, $x_{i1}^{(2)}\xi_{i2}^{(2)}, x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)}, \xi_{i1}^{(2)}\xi_{i1}^{(2)}, \xi_{i2}^{(2)}\xi_{i2}^{(2)}$ in the second group (Kentucky) are quite different from zero, while the quadratic interaction effects corresponding to $\xi_{i1}^{(1)}\xi_{i1}^{(1)}, \xi_{i2}^{(1)}\xi_{i2}^{(1)}$ in first group (Ohio) and $x_{i1}^{(2)}x_{i2}^{(2)}$ in the second group (Kentucky) is small and not significant when given the other linear and interaction effects in the latent variable. As $\xi_{i1}^{(1)}, \xi_{i2}^{(1)}, x_{i1}^{(1)}, x_{i2}^{(1)}$ and $\xi_{i1}^{(2)}, \xi_{i2}^{(2)}, x_{i1}^{(2)}, x_{i2}^{(2)}$ in the first and second groups is related to the mixed ordered categorical and dichotomous variables, the standard error estimate (0.209), (0.250), (0.302) correspond to $\beta_3^{(1)} = (0.189), \beta_4^{(1)} = (-0.251), \beta_5^{(1)} = (0.299)$ is quite large, and the HPD interval, (-0.221, 0.599), (-0.745, 0.238) and (-0.255, 0.956) is quite wide. The standard error estimate (0.275), (0.327), (0.309) and (0.273) correspond to $\beta_4^{(2)} = (-0.176), \beta_5^{(2)} = (0.568)$ and $\gamma_3^{(2)} = (0.272), \gamma_4^{(2)} = (0.373)$ is quite large, and the HPD interval, (-0.697, 0.372), (-0.053, 1.247), (-0.399, 0.839) and (-0.153, 0.914) is quite wide. Hence the corresponding interaction effect is substantial.

The estimated multiple group nonlinear latent variable using truncated normal distribution with equal spaces of thresholds is given by

$$\eta_i^{(1)} = (0.490)x_{i1}^{(1)} + (0.395)x_{i2}^{(1)} + (0.166)x_{i1}^{(1)}x_{i2}^{(1)} + (-0.211)x_{i1}^{(1)}\xi_{i2}^{(1)} + (0.187)x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)} + (0.505)\xi_{i1}^{(1)} + (0.612)\xi_{i2}^{(1)} + (0.001)\xi_{i1}^{(1)}\xi_{i1}^{(1)} + (-0.011)\xi_{i2}^{(1)}\xi_{i2}^{(1)} + \delta_i^{(1)} \quad (25)$$

$$\eta_i^{(2)} = (0.693)x_{i1}^{(2)} + (0.386)x_{i2}^{(2)} + (0.090)x_{i1}^{(2)}x_{i2}^{(2)} + (-0.292)x_{i1}^{(2)}\xi_{i2}^{(2)} + (0.517)x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)} + (0.666)\xi_{i1}^{(2)} + (0.506)\xi_{i2}^{(2)} + (0.131)\xi_{i1}^{(2)}\xi_{i1}^{(2)} + (0.193)\xi_{i2}^{(2)}\xi_{i2}^{(2)} + \delta_i^{(2)} \quad (26)$$

From the standard error estimates and the HPD intervals, we observe that all the linear effects of the covariates and the exogenous latent variables, and the interaction effects corresponding to $x_{i1}^{(1)}x_{i2}^{(1)}, x_{i1}^{(1)}\xi_{i2}^{(1)}, x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)}$ in the first group (Ohio) are quite different from zero, $x_{i1}^{(2)}\xi_{i2}^{(2)}, x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)}, \xi_{i1}^{(2)}\xi_{i1}^{(2)}, \xi_{i2}^{(2)}\xi_{i2}^{(2)}$ in the second group (Kentucky) are quite different from zero, while the quadratic interaction effects corresponding to $\xi_{i1}^{(1)}\xi_{i1}^{(1)}, \xi_{i2}^{(1)}\xi_{i2}^{(1)}$ in first group (Ohio) and $x_{i1}^{(2)}x_{i2}^{(2)}$ in the second group (Kentucky) is small and not significant when given the other linear and interaction effects in the latent variable. Hence the corresponding interaction effect is substantial.

The estimated multiple group nonlinear latent variable using truncated normal distribution with unequal spaces of thresholds is given by

$$\eta_i^{(1)} = (0.453)x_{i1}^{(1)} + (0.428)x_{i2}^{(1)} + (0.168)x_{i1}^{(1)}x_{i2}^{(1)} + (-0.180)x_{i1}^{(1)}\xi_{i2}^{(1)} + (0.299)x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)} + (0.570)\xi_{i1}^{(1)} + (0.662)\xi_{i2}^{(1)} + (-0.029)\xi_{i1}^{(1)}\xi_{i1}^{(1)} + (0.073)\xi_{i2}^{(1)}\xi_{i2}^{(1)} + \delta_i^{(1)} \quad (27)$$

$$\eta_i^{(2)} = (0.761)x_{i1}^{(2)} + (0.376)x_{i2}^{(2)} + (0.073)x_{i1}^{(2)}x_{i2}^{(2)} + (-0.103)x_{i1}^{(2)}\xi_{i2}^{(2)} + (0.448)x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)} + (0.744)\xi_{i1}^{(2)} + (0.517)\xi_{i2}^{(2)} + (0.205)\xi_{i1}^{(2)}\xi_{i1}^{(2)} + (0.291)\xi_{i2}^{(2)}\xi_{i2}^{(2)} + \delta_i^{(2)} \quad (28)$$

From the standard error estimates and the HPD intervals, we observe that all the linear effects of the covariates and the exogenous latent variables, and the interaction effects corresponding to $x_{i1}^{(1)}x_{i2}^{(1)}, x_{i1}^{(1)}\xi_{i2}^{(1)}, x_{i1}^{(1)}\xi_{i1}^{(1)}\xi_{i2}^{(1)}$ in the first group (Ohio) are quite different from zero, $x_{i1}^{(2)}\xi_{i2}^{(2)}, x_{i1}^{(2)}\xi_{i1}^{(2)}\xi_{i2}^{(2)}, \xi_{i1}^{(2)}\xi_{i1}^{(2)}, \xi_{i2}^{(2)}\xi_{i2}^{(2)}$ in the second group (Kentucky) are quite different from zero, while the quadratic interaction effects corresponding to $\xi_{i1}^{(1)}\xi_{i1}^{(1)}, \xi_{i2}^{(1)}\xi_{i2}^{(1)}$ in first group (Ohio) and $x_{i1}^{(2)}x_{i2}^{(2)}$ in the second group (Kentucky) is small and not significant when given the other linear and interaction effects in the latent variable. Hence the corresponding interaction effect is substantial.

The results corresponding to the first and second groups and ordered categorical variables and covariates with hidden continuous normal distribution (censored normal distribution) for variables and hidden continuous normal distribution (truncated normal distribution with known parameters) as well as two types of thresholds (with equal and unequal distances for categories) are reported in Tables (1:2). We noticed that the SE values are small values in the first and second groups.

The results corresponding to the first and second groups and ordered categorical variables and covariates with hidden continuous normal distribution (truncated normal distribution with known parameters) for variables and hidden continuous normal distribution (truncated normal distribution with known parameters) as well as two types of thresholds (with equal and unequal spaces for categories) are reported in Tables (3:4). We observed that the SE values are small values in the first and second groups.

The highest posterior density (HPD) of all the parameters was calculated. We noticed that the effectiveness of the HPD intervals is convenient for ordered categorical variables when using censored normal distribution, truncated normal distribution.

To reveal the effectiveness of DIC for model comparison, we reanalysed the data sets via a nonlinear latent variable model with interaction term in the latent variable model (M_4). The DIC values obtained were compared to those obtained under the correct model. The results are presented in Tables (5:6).

The DIC values of censored normal distribution, truncated normal distribution with equal spaces of thresholds, are (12671) and (13456) respectively.

The model fitting DIC of LVMs with mixed data using censored normal distribution and is less than the model fitting DIC of LVMs with mixed data truncated normal distribution. However, it performs very well for the mixed variables with censored normal distribution.

The DIC values of censored normal distribution, truncated normal distribution with unequal spaces of thresholds, are (12408) and (13467) respectively.

The model fitting DIC of LVMs with mixed data using censored normal distribution and is less than the model fitting DIC of LVMs with mixed data using truncated normal distribution with unequal spaces of thresholds. However, it performs very well for the mixed variables with censored normal distribution.

The best fitted model with a smallest DIC value is a censored normal distribution with unequal spaces of thresholds (12408). Also, the DIC value of truncated normal distribution with equal spaces of thresholds is (13456). As a result, we observed that the performance of DIC is not satisfactory and would be worse with mixed data using truncated normal distribution with unequal spaces of thresholds.

7. Conclusions and Recommendations

The multi-sample nonlinear models which involving nonlinear effects with nonlinear covariates and latent variables are very common in social and behavioural sciences. The first purpose of this analysis was to use multi-sample nonlinear LVMs with nonlinear covariates and latent variables to obtain all the estimated parameters. The second purpose is to solve the problem of mixed variables by using hidden continuous normal distribution (censored normal distribution and truncated normal distribution) and to solve the problem of mixed ordered categorical and dichotomous covariates using hidden continuous normal distribution (truncated normal distribution with known parameters). The proposed procedures have been done using two types of thresholds (with equal and unequal categories distances). However, this assumption is likely to be violated in many practical applications.

In this paper, a Bayesian approach is proposed for analysing multi-sample nonlinear models with mixed 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 mixed data manifest variables are alleviated by data augmentation with some MCMC methods. 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. Booth, B. M., Leukefeld, C., Falck, R., Wang, J. and Carlson, R. (2006). Correlates of rural methamphetamine and cocaine users: Results from a multistate community study. *Journal of Studies on Alcohol and Drugs*, 67(4), 493.
3. Broemeling, L. D. (1985). *Bayesian analysis of linear models*: Dekker New York.
4. Cai, J.-H., Song, X.-Y. and Lee, S.-Y. (2008). Bayesian analysis of non-linear structural equation models with mixed continuous, ordered and unordered categorical, and non-ignorable 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. Lee, S.Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. Chichester, England; Hoboken, NJ: Wiley:.

8. Lee, S.Y. and Song, X.Y. (2012). Basic and advanced structural equation models for medical and behavioural sciences. Hoboken: Wiley.
9. Lee, S.Y. (2006). Bayesian Analysis of Non-linear Structural Equation Models with Non-ignorable Missing Data. *Psychometrika*, 71(3), 541-564. doi: 10.1007/s11336-006-1177-1
10. 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.
11. Lee, S.-Y. and Song, X.-Y. (2003). Model Comparison of Non-linear Structural Equation Models with Fixed Covariates. *PSYCHOMETRIK*, 68(1), 27-47.
12. 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.
13. Rubin, D. B. (1991). EM and beyond. *Psychometrika*, 56(2), 241-254.
14. Song, X.Y. and Lee, S.Y. (2006b). Bayesian analysis of structural equation models with non-linear covariates and latent variables. *Multivariate Behavioral Research*, 41(3), 337-365 .
15. Song, X. Y. and Lee, S.Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical and Statistical Psychology*, 54(2), 237-263.
16. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and & 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.
17. Thanoon, T. Y., and Adnan, R. (2015). Bayesian Analysis of Multiple Group Non-linear Structural Equation Models with Ordered Categorical and Dichotomous Variables: A Survey. *Research Journal of Mathematical and Statistical Sciences*, ISSN, 2320, 6047.
18. Thanoon, T. Y., and Adnan, R. (2016a). Bayesian Analysis of Linear and Non-linear Latent Variable Models with Fixed Covariate and Ordered Categorical Data. *Pakistan Journal of Statistics and Operation Research*, 12(1), 125-140.
19. Thanoon, T. Y., and Adnan, R. (2016c). Comparison between Bayesian structural equation models with ordered categorical data. In *AIP Conference Proceedings* (Vol. 1750, No. 1, p. 060026). AIP Publishing.
20. Thanoon, T. Y., and Adnan, R. (2017). Model comparison of linear and non-linear Bayesian structural equation models with dichotomous data. *Communications in Statistics-Simulation and Computation*, 46(6), 4578-4599.
21. Thanoon, T. Y., Adnan, R., and 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.
22. Thanoon, T. Y., and Adnan, R. (2016b). Bayesian Multi-Sample Non-linear Structural Equation Models with Mixed Ordered Categorical and Dichotomous Variables. *Journal of Applied Probability*, 11(2), 1-16.
23. Thanoon, T. Y., and Adnan, R. (2019). Improve the Bayesian generalized latent variable models with non-linear variable and covariate of dichotomous data. In *AIP Conference Proceedings* (Vol. 2096, No. 1, p. 020009). AIP Publishing.