Statistical analysis of nonlinear structural equation models with continuous and polytomous data

Sik-Yum Lee* and Hong-Tu Zhu

Department of Statistics, The Chinese University of Hong Kong, Hong Kong

A general nonlinear structural equation model with mixed continuous and polytomous variables is analysed. A Bayesian approach is proposed to estimate simultaneously the thresholds, the structural parameters and the latent variables. To solve the computational difficulties involved in the posterior analysis, a hybrid Markov chain Monte Carlo method that combines the Gibbs sampler and the Metropolis–Hasting algorithm is implemented to produce the Bayesian solution. Statistical inferences, which involve estimation of parameters and their standard errors, residuals and outliers analyses, and goodness-of-fit statistics for testing the posited model, are discussed. The proposed procedure is illustrated by a simulation study and a real example.

1. Introduction

Structural equation modelling (SEM) is an important multivariate method in analysing relationships among manifest and latent variables, and in establishing useful models for achieving correct decisions and conclusions. With the availability of efficient software, such as EQS (Bentler, 1992) and LISREL (Jöreskog & Sörbom, 1996), it represents a widely used multivariate method in behavioural, educational and social sciences. At present, most statistical theory and computer software in the field are based on models that involve only linear relationships among the manifest and the latent variables. 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. For example, the theory of reasoned action or expectancy-value attitude theory in social psychology (Ajzen & Fishbein, 1980) involves interactions among theoretical constructs; see Bagozzi (1986) for further explanation and Bagozzi, Baumgartner & Yi (1992) and references therein for its interpretation and application in consumer research. Numerous theories in educational psychology have been proposed to illustrate that combinations of a teaching method with a certain type of reinforcement produce

* Requests for reprints should be addressed to Professor S. Y. Lee, Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, (e-mail: sylee@sparc2.sta.cuhk.edu.hk).
achievement levels that are different from that are expected on the basis of their separate effects. There are more examples of the significance of nonlinear effects in Schumacker & Marcoulides (1998).

Clearly, analysis of nonlinear relationships in SEM is a non-trivial problem. The nonlinear factor analysis model with polynomial relationships was first explored by McDonald (1962, 1967), and followed up by McDonald (1979), Etezadi-Amoli & McDonald (1983) and Mooijaart and Bentler (1986). Recently, methods that use LISREL (Jöreskog & Sörbom, 1996) have been proposed to analyse some simple nonlinear models; for example, see Busemeyer and Jones (1983), Kenny and Judd (1984), Jöreskog and Yang (1996) and Ping (1996), among others. Bollen (1995) and Bollen and Paxton (1998) give a two-stage least squares approach. A comparison of these recent methods has been described by Li et al. (1998). However, rigorous statistical properties of these estimation methods, such as the standard errors estimates and goodness-of-fit statistics, have not been established.

More statistically sound methods for nonlinear factor analysis and SEM have been proposed by Zhu & Lee (1999) and Arminger & Muthén (1998), respectively. Both papers use a Bayesian approach with a hybrid algorithm that combines the Gibbs sampler (Geman & Geman, 1984) and the Metropolis–Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) in producing the estimates. In addition to the Bayesian estimates of the structural parameters, standard error estimates and direct estimates of the latent variables can be obtained. Moreover, a goodness-of-fit statistic for evaluating the posited nonlinear model is given in Zhu & Lee (1999).

All the methods cited above have been developed with the assumption that the available data are multivariate normal. However, in practical applications, many variables in behavioural, educational and social sciences are polytomous. Olsson (1979) and Lee, Poon & Bentler (1992) pointed out that treating the polytomous data as continuous and ignoring their polytomous nature can lead to erroneous conclusions. Clearly, rigorous statistical analysis of SEM with a mixture of continuous and polytomous variables is much more difficult. Reboussin & Liang (1998) proposed a quadratic estimating equation approach for the LISCOMP (Muthén, 1987) model, while Lee, Poon & Bentler (1995) proposed a two-stage procedure to analyse linear SEM with arbitrary covariance structures. Shi & Lee (1998) introduced a Bayesian approach for the linear factor analysis model, in which the structural parameters and the factor scores are simultaneously estimated via the Gibbs sampler (Geman & Geman, 1984). They showed that the major computational burden in evaluating the multiple integrals associated with the polytomous variables can be reduced via the Bayesian approach with Markov chain Monte Carlo (MCMC) methods.

The main objective of this paper is to propose a Bayesian approach for analysing a general nonlinear SEM with mixed continuous and polytomous variables. The underlying model is an extension of the LISREL model with a nonlinear structural equation that specifies the nonlinear relationships among the latent variables. The main idea in handling the polytomous variables in the Bayesian analysis is to treat the underlying latent continuous measurements as hypothetical missing data and augment them with the observed data in the posterior analysis. Using this approach, the model based on the complete data set becomes one with continuous variables; hence the Bayesian solution can be obtained via a hybrid algorithm that combines the Gibbs sampler (Geman & Geman, 1984) and the MH algorithm (Metropolis et al., 1953; Hastings, 1970). Simultaneous estimates of the unknown thresholds, structural parameters and latent variables are produced. In addition to these point estimates, we also
develop statistical methods to obtain standard errors estimates, methods to analyse residuals and outliers, and statistics for assessing the goodness of fit of the posited model.

The paper is organized as follows. A general nonlinear SEM with mixed continuous and polytomous variables is described in Section 2. Section 3 contains the main theoretical developments. Here, the Bayesian approach and the hybrid algorithm are described; and the conditional distributions for implementation of the hybrid algorithm are derived. Statistical analyses based on the observations simulated by the hybrid algorithm are discussed in Section 4. Empirical results, which are obtained from a simulation study and a real example, are presented in Section 5. Some concluding remarks are given in Section 6. In the following sections, \( p(\cdot), p(\cdot|\cdot) \) and \([\cdot|\cdot]\) denote the density function, the conditional density function, and the conditional distribution, respectively; and \( I_q \) denotes the identity matrix of order \( q \).

## 2. Model description

Consider the following factor analytic measurement model for the \( p \times 1 \) manifest random vector \( U_i \):

\[
U_i = \mu + \Lambda \xi_i + \epsilon_i, \quad i = 1, \ldots, n,
\]

where \( \mu \) is the \((p \times 1)\) mean vector, \( \Lambda \) is the \((p \times q)\) factor loading matrix, \( \xi_i \) is a \((q \times 1)\) latent random vector and \( \epsilon_i \) is a \((p \times 1)\) random vector of error measurements with distribution \( N[0, \Psi_\epsilon] \), where \( \Psi_\epsilon \) is diagonal and \( \epsilon_i \) is independent of \( \xi_i \). To handle more complex situations, we partition \( \xi_i \) as \((\xi_i^{(1)}, \xi_i^{(2)})'\) and further model this latent vector via the following nonlinear structural model:

\[
\xi_i^{(1)} = \Pi \xi_i^{(1)} + \Gamma H(\xi_i^{(2)}) + \delta_i,
\]

where \( \xi_i^{(1)} \) (\( q_1 \times 1 \)) and \( \xi_i^{(2)} \) (\( q_2 \times 1 \)) are latent subvectors of \( \xi_i \); \( H(\xi_i^{(2)}) = (h_1(\xi_i^{(2)}), \ldots, h_t(\xi_i^{(2)}))' \) is a \( t \times 1 \) vector-valued function with differentiable functions \( h_1, \ldots, h_t \), and \( t \geq q_2 \); \( \Pi \) (\( q_1 \times q_1 \)) and \( \Gamma \) (\( q_1 \times t \)) are matrices of regression coefficients on \( \xi_i^{(1)} \) and \( H(\xi_i^{(2)}) \), respectively; \( \xi_i^{(2)} \) and \( \delta_i \) are independently distributed as \( N[0, \Phi_\xi] \) and \( N[0, \Phi_\delta] \), where \( \Phi_\delta \) is a diagonal covariance matrix. It is assumed that \( |I_q - \Pi| \neq 0 \). The structural model in (2) is linear in the parameter matrices \( \Pi \) and \( \Gamma \), but is nonlinear in the latent variables in \( \xi_i^{(1)} \). Hence, nonlinear causal effects of latent variables in \( \xi_i^{(2)} \) on latent variables in \( \xi_i^{(1)} \) can be assessed. Let \( \Lambda_\xi = (\Pi, \Gamma) \) and \( G(\xi) = (\xi^{(1)}, H(\xi^{(2)}))' \); then equation (2) can be written as

\[
\xi_i^{(1)} = \Lambda_\xi G(\xi_i) + \delta_i.
\]

Let \( U = \{X, Y\} \), where \( X = \{x_1, \ldots, x_r\} \), is a subset of variables whose exact continuous measurements are observable, while \( Y = \{y_1, \ldots, y_s\} \) is the remaining subset of variables such that \( p \geq s = p - r \geq 0 \), the corresponding continuous measurements are unobservable, and their information is given by an observable polytomous vector \( Z = (z_1, \ldots, z_s)' \). The variables in \( X \) or \( Y \) can be indicators for latent variables in \( \xi^{(1)} \) and/or \( \xi^{(2)} \). The relationship between \( Y \) and \( Z \) is defined by a set of unknown thresholds as follows:

\[
Z = \begin{bmatrix} z_1 \\ \vdots \\ z_s \end{bmatrix} \quad \text{if} \quad \begin{cases} \alpha_{1,z_1} < y_1 \leq \alpha_{1,z_1+1}, \\ \vdots \\ \alpha_{s, z_s} < y_s \leq \alpha_{s, z_s+1}, \end{cases}
\]
where $z_k$ is an integral value in $\{0, 1, \ldots, b_k\}$ and $\alpha_k = \{\alpha_{k,1}, \ldots, \alpha_{k,b_k}\}$. In general, we set $\alpha_{k,0} = -\infty, \alpha_{k,b_k} = \infty$. For the $k$th variable, there are $b_k + 1$ categories which are defined by the unknown thresholds $\alpha_{k,j}$. Suppose that the model relating to the subvector $Y_i = (y_{1i}, \ldots, y_{ni})'$ of $U_i$ is given by

$$Y_i = \mu_y + \Lambda_y \xi_i + \epsilon_{yi}$$

where $\mu_y$ is an $(s \times 1)$ subvector of $\mu$, $\Lambda_y$ is an $(s \times q)$ submatrix of $\Lambda$, $\epsilon_{yi}$ is an $(s \times 1)$ subvector of $\epsilon_i$ with diagonal covariance matrix $\Psi_{yi}$. Let $Z_i = (z_{1i}, \ldots, z_{ni})'$ be the $i$th observation of $Z$ corresponding to $Y_i$, $i = 1, \ldots, n$. In this paper, the nonlinear SEM defined by equations (1) and (2) will be analysed with the data $\{(X_i, Z_i), i = 1, \ldots, n\}$ of mixed continuous and polytomous observations. It can be seen that our basic model is similar to that in Arminger & Muthén (1998); however, their development cannot be applied to polytomous data. Our model also includes certain important special cases. For example, let $\Lambda = (\Lambda'_1, \Lambda'_2)$ be an appropriate partition corresponding to $\xi^{(1)}$ and $\xi^{(2)}$; if $H(\xi^{(2)}) = \xi^{(2)}$, it reduces to the ordinary LISREL model with polytomous variables. Hence, the factor analysis model with polytomous variables considered in Shi & Lee (1998) can also be regarded as a special case.

It has been pointed out by Lee, Poon & Bentler (1989) that the mean, the variance and the thresholds of a polytomous variable are not identified. Hence, some identification conditions on the parameters associated with such variables are required. In particular, the model is identified by imposing the following restrictions: $\mu_y = 0$, and $\text{diag}\{\text{Var}(Y)\} = I_s$. Another method is to impose appropriate constraints on the mean and the thresholds. For example, (a) fix any pair $(\alpha_{k,j_1}, \alpha_{k,j_2})$ for each $k = 1, \ldots, s$ and $j_1 \neq j_2$; or (b) fix any $\alpha_{k,j}$ for each $k = 1, \ldots, s$, and fix the corresponding mean vector to be a preassigned constant vector. The identification condition, $\text{diag}\{\text{Var}(Y)\} = I_s$, imposes nonlinear restrictions on the parameters, thus the development of the theory and the algorithm in achieving the Bayesian solution is more complicated. Moreover, for the nonlinear SEM, it is generally difficult to derive the variance of $Y$. Hence, identification conditions given in (a) with $j_1 = 1$ and $j_2 = b_k$ will be imposed in this paper. In this way, the dispersion of the polytomous variable is controlled by the fixed thresholds. From a Bayesian point of view, it is equivalent to take the prior distribution of the threshold with probability 1 at the corresponding preassigned value. It is straightforward to adapt conditions given in (b) via similar developments. Finally, to identify the covariance structures, some appropriate elements in $\Lambda$ and $\Lambda_\xi$ are set equal to some fixed known values. Of course, other identification conditions, such as those given in Etezadi-Amoli & McDonald (1983), may also be considered.

### 3. Posterior analysis

Let $X = (X_1, \ldots, X_n)$ and $Z = (Z_1, \ldots, Z_n)$ be the observed continuous and polytomous data matrices, respectively. Based on $X$ and $Z$, analysis of the nonlinear SEM as defined in (1) and (2) involves two serious difficulties: one is due to the nonlinearity of $H(\xi^{(2)})$ and the other is due to the polytomous nature of $Z$. Hence, existing methods, such as generalized least squares, maximum likelihood, two-step procedures (see Lee et al., 1995), asymptotically distribution-free method (Bentler; 1983; Browne, 1984), and the Bayesian procedures developed in Arminger & Muthén (1998) and Zhu & Lee (1999) are not applicable. Here,
we propose a Bayesian procedure that uses the following powerful strategy recently developed in statistical computing (see Rubin, 1991; Meng & Van Dyk, 1997): treat the latent continuous measurements \( Y = (Y_1, \ldots, Y_n) \) and the latent variables \( F = (\xi_1, \ldots, \xi_n) \) as missing, and augment the observed data \([X, Z]\) with the hypothetical missing data \([Y, F]\) in the posterior analysis. As we will see later, the above-mentioned difficulties can be solved via this approach. Joint Bayesian estimates of \( F \), the thresholds in \( \alpha = (\alpha_1, \ldots, \alpha_s) \) and the structural parameter vector \( \theta \) that contains all unknown parameters in \( \mu, \Phi_\xi, \Lambda, \Lambda_\xi, \Psi_\delta \) and \( \Psi_\epsilon \), will be obtained.

In a Bayesian framework, the unknown \( \alpha \) and \( \theta \) are regarded as random vectors that come from some appropriate prior distributions. The joint distribution of all the random variables can be written as

\[
p(\alpha, \theta, F, X, Y, Z) = p(\alpha, \theta)p(F|\alpha, \theta)p(X, Y, Z|F, \alpha, \theta).
\]

It follows from the definitions of \( F, X, Y \) and \( Z \) that \( p(F|\alpha, \theta) = p(F|\theta) \) and \( p(X, Y, Z|F, \alpha, \theta) = p(U|F, \theta)p(Z|Y, \alpha) \) where \( U = (U_1, \ldots, U_n) \) with \( U_i = \{X_i, Y_i\} \). It is also natural to assume that \( p(\alpha, \theta) = p(\alpha)p(\theta) \). Hence, the joint distribution of all the random variables is expressed by the following factorization:

\[
p(\alpha, \theta, F, X, Y, Z) = p(\alpha)p(\theta)p(F|\theta)p(U|F, \theta)p(Z|Y, \alpha).
\]

Let \( \theta_\epsilon \) be the unknown parameters in \( \mu, \Lambda \) and \( \Psi_\epsilon \) associated with (1); and \( \theta_\xi \) be the unknown parameters in \( \Lambda_\xi, \Phi_\xi \) and \( \Psi_\delta \) associated with (2). Further, assume that the prior distribution of \( \theta_\epsilon \) is independent of the prior distribution of \( \theta_\xi \), that is, \( p(\theta) = p(\theta_\epsilon)p(\theta_\xi) \); and \( p(F|\theta) = p(F|\theta_\xi) \) and \( p(U|F, \theta) = p(U|F, \theta_\epsilon) \). Hence,

\[
p(\alpha, \theta, F, X, Y, Z) = p(\alpha)p(\theta_\epsilon)p(\theta_\xi)p(F|\theta_\xi)p(U|F, \theta_\epsilon)p(Z|Y, \alpha).
\]

In a Bayesian approach, we need to analyse the posterior distribution \([\alpha, \theta, F|X, Z]\). Due to the polytomous nature of \( Z \) and the nonlinearity of \( H(\xi^{(2)}) \), this posterior distribution involves intractable multiple integrals. Our idea for handling this difficult problem is to augment the observed data \([X, Z]\) further with the latent continuous measurements \( Y \) in the posterior analysis. The Gibbs sampler (Geman & Geman, 1984) will be used in generating a sequence of random observations from the joint posterior distribution \([\alpha, \theta, F, Y|X, Z]\). Then, the Bayesian solution is obtained by means of the generated observations. To implement the Gibbs sampler, one starts with initial values \((\alpha^{(0)}, \theta^{(0)}, F^{(0)}, Y^{(0)})\), then simulates \((\alpha^{(1)}, \theta^{(1)}, F^{(1)}, Y^{(1)})\) and so on; at the \( m \)th iteration, with current values \( \alpha^{(m)}, \theta^{(m)}, F^{(m)} \) and \( Y^{(m)} \):

(a) Generate \( F^{(m+1)} \) from \( p(F|\alpha^{(m)}, \theta^{(m)}, Y^{(m)}, X, Z) \);

(b) Generate \( \theta^{(m+1)} \) from \( p(\theta|\alpha^{(m)}, F^{(m+1)}, Y^{(m)}, X, Z) \);

(c) Generate \( (\alpha^{(m+1)}, Y^{(m+1)}) \) from \( p(\alpha, Y|\theta^{(m+1)}, F^{(m+1)}, X, Z) \).

The cycle defined in (6) is iterated \( j \) times, generating \((\alpha^{(j)}, \theta^{(j)}, F^{(j)}, Y^{(j)})\). As \( j \) approaches infinity, the joint distribution of \((\alpha^{(j)}, \theta^{(j)}, F^{(j)}, Y^{(j)})\) can be shown to approach the joint posterior distribution \([\alpha, \theta, F, Y|X, Z]\); see Geman & Geman (1984) and Geyer (1992). The simulated sequence of random observations from the joint posterior distribution will be used to obtain the Bayesian solution as described in Section 4. The posterior simulations associated with the Gibbs sampler are derived as follows.
3.1. Posterior simulations

Because $\xi$ are conditionally independent and $U_i$ are also conditionally independent, it can be shown that

$$p(F | \alpha, \theta, Y, X, Z) = \prod_{i=1}^{n} p(\xi_i | U_i, \theta) \propto \prod_{i=1}^{n} p(U_i | \xi_i, \theta) p(\xi_i^{(1)} | \xi_i^{(2)}, \theta) p(\xi_i^{(2)} | \theta).$$

Based on the definition of the model and assumptions, $p(\xi_i | U_i, \theta)$ is proportional to

$$\exp \left\{ -\frac{1}{2}(\xi_i^{(2)} - \Phi_{\xi}^{-1}\xi_i^{(2)}) - \frac{1}{2}(U_i - \mu - \Delta_{\xi})'\Psi_{\xi}^{-1}(U_i - \mu - \Delta_{\xi}) - \frac{1}{2}(\xi_i^{(1)} - \Delta_{\xi}G(\xi_i))'\Psi_{\xi}^{-1}(\xi_i^{(1)} - \Delta_{\xi}G(\xi_i)) \right\}. \quad (7)$$

Hence, the conditional distribution required in step (a) is achieved.

To derive the conditional distributions with respect to the structural parameters in step (b), it is important to note that since $F$ and $Y$ are given, the models defined in (1) and (2) reduce to the standard linear models. Hence, the difficulties arising from the nonlinearity of $H(\xi^{(2)})$ and the nature of $Z$ are alleviated significantly. According to the suggestions made in Lindley & Smith (1972), Lee (1981), Broemeling (1985) and Shi & Lee (1998), the following commonly used conjugate-type prior distributions can be applied to situations where we have a rough idea about the hyperparameters:

$$p(\mu) \sim N[\mu_0, \Sigma_0], \quad p(\psi_{ek}^{-1}) \sim Gamma[\alpha_{0,ek}, \beta_{0,ek}],$$

$$p(\Lambda_{ek} | \psi_{ek}) \sim N[\Lambda_{0,ek}, \psi_{ek}H_{0,ek}], \quad k = 1, \ldots, p,$$

where $\Lambda_{ek}$ is a $r_{ek} \times 1$ row vector that contains the unknown parameters in the $k$th row of $\Lambda$; $\alpha_{0,ek}, \beta_{0,ek}, \mu_0, \Lambda_{0,ek}, H_{0,ek}$, and $\Sigma_0$ are hyperparameters whose values are assumed to be given. For $k \neq h$, it is assumed that $(\psi_{ek}, \Lambda_{ek})$ and $(\psi_{eh}, \Lambda_{eh})$ are independent.

To cope with the case of fixed known elements in $\Lambda$, let $C = (c_{kj})$ be the index matrix such that $c_{kj} = 0$ if $\lambda_{kj}$ is known and $c_{kj} = 1$ if $\lambda_{kj}$ is unknown, $r_{ek} = \sum_{i=1}^{r} c_{ki}; F_k$ be a submatrix of $F$ such that the $j$th row with $c_{kj} = 0$ deleted; and $U_{ki}^* = (U_{k1}^*, \ldots, U_{kn}^*)$ with

$$U_{ki}^* = U_{ki} - \mu_k - \sum_{j=1}^{r} \lambda_{kj} \xi_{j} (1 - c_{kj}),$$

where $U_{ki}$ is the $(k + h)$th element of $U_i$ and $\mu_k$ is the $k$th element of $\mu$. Let $\Omega_k = (H_{0,ek}^{-1} + F_kF_k'^{-1})^{-1}, \nu_k = (\xi_{j} - \Delta_{\xi}G(\xi_{j})) - \Delta_{\xi}G(\xi_{j})^{-1}(\xi_{j} - \Delta_{\xi}G(\xi_{j}))$, and $\beta_{ek} = \beta_{0,ek} + 2^{-1} (U_{ki}^* - \nu_k \Omega_k^{-1} \nu_k + \Lambda_{0,ek}H_{0,ek}^{-1} \Lambda_{0,ek})$. Then, it can be shown by similar reasoning as in Shi & Lee (1998) that, for $k = 1, \ldots, p$,

$$p(\gamma_{ek} | U, F, \mu) \sim Gamma[n/2 + \alpha_{0,ek}, \beta_{ek}], \quad p(\Lambda_{ek} | U, F, \gamma_{ek}, \mu) \sim N[\nu_k, \gamma_{ek}^{-1} \Omega_k],$$

$$p(\mu | U, F, \Lambda, \Psi_{e}) \sim N \left[ (\Sigma_0^{-1} + n\Psi_{e}^{-1})^{-1}(n\Psi_{e}^{-1}\tilde{B} + \Sigma_0^{-1}\mu_0), (\Sigma_0^{-1} + n\Psi_{e}^{-1})^{-1} \right], \quad (8)$$

where $\gamma_{ek} = \psi_{ek}^{-1}$ and $\tilde{B} = \sum_{i=1}^{n} (U_i - \Delta_{\xi}i)/n$.

Now, consider the conditional distribution of $\theta$ that is proportional to $p(F | \theta) p(\theta)$. Let $F_{(1)} = (\xi_{1}^{(1)}, \ldots, \xi_{n}^{(1)}), \quad F_{(2)} = (\xi_{1}^{(2)}, \ldots, \xi_{n}^{(2)})$ and $G = (G(\xi_{1}), \ldots, G(\xi_{n}))$. Since the
distribution of $\xi_i^{(2)}$ only involves $\Phi_\xi$, $p(F(2) | \theta_\xi) = p(F(2) | \Phi_\xi)$. Moreover, it is assumed that the prior distribution of $\Phi_\xi$ is independent of the prior distribution of $\Lambda_\xi$ and $\Psi_\delta$. It follows that
\[
p(F | \theta_\xi) p(\theta_\xi) \propto \prod_i p(F(i) | F(2), \Lambda_\xi, \Psi_\delta) p(\Lambda_\xi, \Psi_\delta) | p(F(2) | \Phi_\xi) p(\Phi_\xi).
\]

Hence, the marginal conditional densities of $(\Lambda_\xi, \Psi_\delta)$ and $\Phi_\xi$ can again be treated separately.

Consider a conjugate-type prior distribution for $\Phi_\xi$ with $p(\Phi_\xi^{-1}) \sim W[R_0, \rho_0, q_2]$, where $W[\cdot, \cdot, \cdot]$ denotes the Wishart distribution, and $\rho_0$ and the positive definite matrix $R_0$ are the given hyperparameters. It can be shown (see Shi & Lee, 1998) that
\[
p(\Phi_\xi | F(2)) \sim IW[(F(2)F' + R_0^{-1}), n + \rho_0, q_2].
\]

where $IW[\cdot, \cdot, \cdot]$ denotes the inverted Wishart distribution.

As before, the prior distributions of elements in $(\Psi_\delta, \Gamma)$ are taken as
\[
p(\psi_\delta^{-1}) \sim Gamma[\alpha_0, \beta_0], p(\Gamma_k | \psi_\delta) \sim N[\Gamma_0, \psi_\delta H_0].
\]

where $k = 1, \ldots, q_1$, $\Gamma_k$ is a $r_{y_{\delta}} \times 1$ row vector that contains the unknown parameters in the $k$th row of $\Gamma$; and $\alpha_0, \beta_0, \beta_0, \sum_0, \Gamma_0, H_0$ and $H_0$ are all given hyperparameters. For $h \neq k$, $(\psi_\delta, \Gamma_k)$ are assumed to be independent.

Let $C_\xi = (c_{\xi kj})$ be the index matrix associated with $\Gamma$, $G_k^*$ be the submatrix of $G$ such that all the $j$th row elements corresponding to $c_{\xi kj} = 0$ were deleted; and $F_k^* = (\xi_{k1}^{(1)}, \ldots, \xi_{kn}^{(1)*})$ with
\[
\xi_{ki}^{(1)*} = \xi_{ki}^{(1)} - \sum_{j=1, k_{\xi_{ki}} = 0}^{q_1} \pi_{kj} \xi_{ji}^{(1)} - \sum_{j=1}^{t} \gamma_{kj} \delta H_j \xi_i^{(2)} (1 - c_{\xi kj}).
\]

Then, it can be shown (see Shi & Lee, 1998) that for $k = 1, \ldots, q_1$,
\[
p(\gamma_\delta | Z, \Pi) \sim Gamma[n/2 + \alpha_0, \beta_0], p(\Gamma_k | Z, \gamma_\delta, \Pi) \sim N[\nu_\delta, \gamma_\delta^{-1} \Omega_\delta^{-1}]
\]

where $\gamma_\delta = \psi_\delta^{-1}$, $\Omega_\delta = (H^{-1}_0 + G_k^* G_k^{*-1})^{-1}$, $\nu_\delta = \Omega_\delta^{-1} \nu_\delta H_0^{-1} \Gamma_0^{-1}$, and $\beta_\delta = \beta_0$ +
\[2^{-1}(F_k^* F_k^* - \nu_\delta^{-1} \Omega_\delta^{-1} \nu_\delta + \Gamma_k^* H_k^{-1} \Gamma_k).
\]

Let $\Pi u$ be a $q_{\pi} \times 1$ vector that contains unknown parameters in $\Pi$, and with prior distribution $N(\Pi u_0, \Sigma_{\pi 0})$. Then, $p(\Pi u | \cdot)$ is proportional to
\[
\exp \left\{ n \log |I_{\Pi 1} - \Pi| - \frac{1}{2}(\Pi u - \Pi u_0)' \Sigma_{\pi 0}^{-1} (\Pi u - \Pi u_0) \\
- \frac{1}{2} \sum_{i=1}^{n} (\xi_i^{(1)} - \Lambda_{\xi} G(\xi_i))' \Psi_\delta^{-1} (\xi_i^{(1)} - \Lambda_{\xi} G(\xi_i)) \right\}.
\]

This is essentially a normal distribution if $|I_{\Pi 1} - \Pi|$ is independent of $\Pi$.

Finally, we consider the joint conditional distribution of $(\alpha, Y)$ given $\theta, F, X$ and $Z$. To deal with the general situation with little or no information about the values of the thresholds, the following non-informative prior distribution is used:
\[
p(\alpha_k) = p(\alpha_k, 2, \ldots, \alpha_k, h_k-1) \propto c, \text{ for } \alpha_{k,2} < \ldots < \alpha_{k, h_k-1}, \quad k = 1, \ldots, s,
\]

where $c$ is a constant. Given the structure parameter $\theta$ and that the covariance matrix $\Psi$ is diagonal, the polychotomous data $Y$ and the thresholds corresponding to different rows are also conditionally independent. For $k = 1, \ldots, p$, let $Y_k$ and $Z_k$ be the $k$th rows of $Y$ and $Z$, respectively.
The efficiency of the Gibbs sampler algorithm depends heavily on how easily one can sample observations from the conditional distributions. Let

\[ p(\alpha_k, Y_k | Z_k, \theta, F) = p(\alpha_k | Z_k, \theta, F) p(Y_k | \alpha_k, Z_k, \theta, F), \]

with

\[ p(\alpha_k | Z_k, \theta, F) \propto \prod_{i=1}^{n} \left[ \Phi(\psi^{-1/2}_y [\alpha_{k,z_{k,i+1}} - \mu_{yk} - \Lambda_{yk}^t \xi_i]) - \Phi(\psi^{-1/2}_y [\alpha_{k,z_{k,i}} - \mu_{yk} - \Lambda_{yk}^t \xi_i]) \right], \]

and

\[ p(y_{ki} | \alpha_k, Z_k, \theta, F) \sim N(\mu_{yk} + \Lambda_{yk}^t \xi_i, \psi_{yk}) I_{(\alpha_{k,z_{k,i}}, \alpha_{k,z_{k,i+1}})}(y_{ki}), \]

where \( \psi_{yk} \) is the \( k \)th diagonal element of \( \Psi_{yk} \), \( \mu_{yk} \) is the \( k \)th element of \( \mu_y \), \( \Lambda_{yk} \) is the \( k \)th row of \( \Lambda_y \), \( \phi(\cdot) \) is the standard normal density, \( I_A(y) \) is an index function which takes 1 if \( y \in A \) and 0 otherwise, and \( \Phi(\cdot) \) denotes the standard normal cumulative distribution function. As a result,

\[ p(\alpha_k, Y_k | Z_k, \theta, F) \propto \prod_{i=1}^{n} \Phi(\psi^{-1/2}_y [Y_{ki} - \mu_{yk} - \Lambda_{yk}^t \xi_i]) I_{(\alpha_{k,z_{k,i}}, \alpha_{k,z_{k,i+1}})}(y_{ki}). \]

In the Gibbs sampler given in Shi & Lee (1998), \( \alpha \) and \( Y \) were simulated separately from the corresponding marginal conditional distributions. According to Liu (1994), drawing \( (\alpha, Y) \) from the joint conditional distribution as proposed here is more efficient.

Non-informative prior distributions (see Zellner, 1971) can be also considered in deriving the conditional distributions associated with the structural parameters in \( \theta \). It can be shown by similar reasoning to that employed above that these conditional distributions can be obtained by setting the hyperparameters \( \alpha_{\theta ek} \rightarrow -r_{yk}/2, \alpha_{\theta 0k} \rightarrow -r_{\xi k}/2, \beta_{\theta ek} \rightarrow 0, \beta_{\theta 0k} \rightarrow 0, H_{\theta ek}^{-1} \rightarrow 0, H_{\theta 0k}^{-1} \rightarrow 0, R_{\theta}^{-1} \rightarrow 0 \) and \( \Sigma_0^{-1} \rightarrow 0 \) in the appropriate expressions.

### 3.2 Implementation

The efficiency of the Gibbs sampler algorithm depends heavily on how easily one can sample observations from the conditional distributions. It can be seen that the conditional distributions associated with (8), (9) and (10) in step (b) are familiar, and drawing observations from them is straightforward and fast. Since the numbers of latent vectors in \( F \), the latent continuous measurements \( Y \) and the thresholds \( \alpha \) are larger than the number of structural parameters, the simulation in steps (a) and (c) plays a more important role in the algorithm. From (7), (12) and (13), we see that it is difficult to sample from \( p(\xi_i | U_i, \theta) \) and \( p(\alpha_k, Y_k | \theta, F, X, Z) \), which are non-standard and complex. Hence, a hybrid algorithm is required to simulate efficiently from these conditional distributions.

The Metropolis–Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) is a well-known MCMC method that has been widely used to simulate observations from a target density via the help of a proposal distribution from which it is easy to sample. Inspired by the work of Arminger & Muthén (1998), and Zhu & Lee (1999), this algorithm is applied here to generate observations from our target densities \( p(\xi_i | U_i, \theta) \) and \( p(\alpha_k, Y_k | \theta, F, X, Z) \).

For \( p(\xi_i | U_i, \theta) \), we choose \( N[0, \sigma^2 \Omega] \) as the proposal distribution, where \( \Omega^{-1} = \Sigma^{-1}_\xi + \Delta \Psi^{-1}_e \Delta \), with

\[
\Sigma^{-1}_\xi = \begin{bmatrix}
\Pi_0 \Psi^{-1}_e \Pi_0, & -\Pi_0 \Psi^{-1}_e \Gamma \Delta \\
-\Delta \Gamma^t \Psi^{-1}_e \Pi_0, & \Phi^{-1}_\xi + \Delta \Gamma^t \Psi^{-1}_e \Gamma \Delta
\end{bmatrix},
\]

and
where \( \mathbf{I}_{0} = \mathbf{I}_{q1} - \mathbf{I} \) and \( \Delta = \partial H(\xi_{2i})/\partial \xi_{2i} \big|_{\xi_{2i}=0} \). Let \( p(\cdot | 0, \sigma, \Omega) \) be the proposal density corresponding to \( N[0, \sigma^{2}\Omega] \). The MH algorithm for our problem is implemented as follows. At the \( m \)th iteration with a current value \( \xi_{i}^{(m)} \), a new candidate \( \xi_{i} \) is generated from \( p(\cdot | \xi_{i}^{(m)}, \sigma, \Omega) \) and accepted with probability

\[
\min \left\{ 1, \frac{p(\xi_{i} | U_{i}, \Theta)}{p(\xi_{i}^{(m)} | U_{i}, \Theta)} \right\}.
\]

The variance \( \sigma^{2} \) can be chosen such that the average acceptance rate is approximately 0.25 or more, see Gelman, Roberts & Gilks (1995). Hence, a random-walk Metropolis–Hastings algorithm is used within the Gibbs sampler to simulate observations from \( p(\xi_{i} | U_{i}, \Theta) \). Other alternatives such as the ‘independence sampler’ or ‘Langevin–Hastings’ algorithms can also be used. To obtain a good initial value of \( \xi_{i} \), we generate \( \xi_{i} \) from \( p(\cdot | U_{i}, 1, \Omega) \) at the first several iterations, where \( u_{i} = \Omega[(\mathbf{I}_{0}, -\Gamma \Delta)^{-1} \mathbf{H}(\Theta) + \Delta \psi^{-1}(U_{i} - \mu)] \). Strategies for improving the efficiency of the above algorithm are also possible by using a random proposal distribution. This idea was first suggested by Besag, Green, Higdon and Mengersen (1995). For example, we may choose two different proposal distributions \( N[0, \sigma_{1}^{2}\Omega] \) and \( N[0, \sigma_{2}^{2}\mathbf{I}_{q}] \) with preassigned probabilities \( p_{1} \) and \( 1 - p_{1} \). This will increase the speed through the target function and decrease the risk of multiple maxima. The choice of a small \( \sigma_{2} \) will increase the frequency of acceptance.

In the MH algorithm for generating observations from \( p(\alpha_{k}, Y_{k} | \Theta, F, X, Z) \), a suitable joint proposal density for \( \alpha_{k} \) and \( Y_{k} \) is constructed according to the following factorization as in Cowles (1996):

\[
p(\alpha_{k}, Y_{k} | Z_{k}, \Theta, F) = p(\alpha_{k})p(Y_{k} | \alpha_{k}, Z_{k}, \Theta, F).
\]

We generate a vector of thresholds \( (\alpha_{k2}, \ldots, \alpha_{kBk - 1}) \) from the following truncated normal distribution

\[
\alpha_{kz} \sim N(\alpha_{kz}^{(m)}, \sigma_{\alpha_k}^{2})I_{(\alpha_{kz - 1}, \alpha_{kz + 1})}, (\alpha_{kz}) \text{ for } z = 2, \ldots, Bk - 1,
\]

where \( \alpha_{kz}^{(m)} \) is the value of \( \alpha_{kz} \) at the \( m \)th iteration of the Gibbs sampler, and \( \sigma_{\alpha_k}^{2} \) is an appropriate preassigned constant. It follows from the MH algorithm that the acceptance probability for \( (\alpha_{k}, Y_{k}) \) as a new observation is \( \min\{1, R_{k}\} \), where

\[
R_{k} = \frac{p(\alpha_{k}, Y_{k} | \Theta, F, Z_{k})p(\alpha_{k}^{(m)}, Y_{k}^{(m)} | \alpha_{k}, Y_{k}, Z_{k}, \Theta, F)}{p(\alpha_{k}^{(m)}, Y_{k}^{(m)} | \Theta, F, Z_{k})p(\alpha_{k}, Y_{k} | \alpha_{k}^{(m)}, Y_{k}^{(m)}, Z_{k}, \Theta, F)}.
\]

It can be shown from (13), (16) and (17) that

\[
R_{k} = \prod_{z=2}^{B_k-1} \frac{\Phi((\alpha_{kz+1}^{(m)} - \alpha_{kz}^{(m)})/\sigma_{\alpha_k}) - \Phi((\alpha_{kz-1}^{(m)} - \alpha_{kz}^{(m)})/\sigma_{\alpha_k})}{\Phi((\alpha_{kz+1}^{(m)} - \alpha_{kz}^{(m)})/\sigma_{\alpha_k}) - \Phi((\alpha_{kz-1}^{(m)} - \alpha_{kz}^{(m)})/\sigma_{\alpha_k})} \times \prod_{i=1}^{n} \frac{\Phi(\psi_{yk}^{-1/2}[\alpha_{kz,iz}^{(m)} - \mu_{yk} - \Lambda_{yk,i}^{'}(\xi_{i})]) \Phi(\psi_{yk}^{-1/2}[\alpha_{kz,iz}^{(m)} - \mu_{yk} - \Lambda_{yk,i}^{'}(\xi_{i})])}{\Phi(\psi_{yk}^{-1/2}[\alpha_{kz,iz}^{(m)} - \mu_{yk} - \Lambda_{yk,i}^{'}(\xi_{i})]) \Phi(\psi_{yk}^{-1/2}[\alpha_{kz,iz}^{(m)} - \mu_{yk} - \Lambda_{yk,i}^{'}(\xi_{i})])}.
\]

Since \( R_{k} \) depends only on the old and new values of \( \alpha_{k} \) and not on the \( Y_{k} \), it does not need to generate a new \( Y_{k} \) in any iteration in which the new value of \( \alpha_{k} \) is not accepted (see Cowles, 1996). Random observations from the truncated normal in (17) are simulated via the
algorithm of Roberts (1995). Finally, observations from a general $p(\Pi|\cdot)$ can be simulated via the Langevin–Hastings algorithm; see Roberts & Rosenthal (1998).

4. Statistical analyses

4.1. Bayesian estimation

It has been shown (Geman & Geman, 1984; Geyer, 1992) that under mild conditions and for sufficiently large $j$, say $J$, the joint distribution of $(\alpha^{(J)}, \theta^{(J)}, Y^{(J)}, F^{(J)})$ converges at an exponential rate to the desired posterior distribution $[\alpha, \theta, F, Y|X, Z]$. Hence, $[\alpha, \theta, F, Y|X, Z]$ can be approximated by the empirical distribution of $\{(\alpha^{(m)}, \theta^{(m)}, F^{(m)}, Y^{(m)}): m = J + 1, \ldots, J + M\}$ where $M$ is chosen to give sufficient precision to the empirical distribution. The convergence of the algorithm is monitored by the ‘estimated potential scale reduction (EPSR)’ values suggested by Gelman & Rubin (1992). To obtain a more nearly independent sample, observations may be collected in cycles with indices $m = J + s^*, J + 2s^*, \ldots, J + Ms^*$ for some spacing $s^*$ (see Gelfand & Smith, 1990). However, in most practical applications a small $s^*$ will suffice for many statistical analyses such as obtaining estimates of the parameters and standard errors; see Zeger & Karim (1991), Albert & Chib (1993) and Shi & Lee (1998).

For brevity, let $\{(\alpha^{(m)}, \theta^{(m)}, F^{(m)}, Y^{(m)}): m = 1, \ldots, M\}$ be the random observations of $(\alpha, \theta, F, Y)$ generated from $[\alpha, \theta, F, Y|X, Z]$ by the proposed hybrid algorithm. The Bayesian estimates of $\alpha$ and $\theta$, and a direct Bayesian estimate of $F$, can be obtained easily via the corresponding sample means of the generated observations as follows:

$$\hat{\alpha} = M^{-1} \sum_{m=1}^{M} \alpha^{(m)}, \quad \hat{\theta} = M^{-1} \sum_{m=1}^{M} \theta^{(m)}, \quad \hat{F} = M^{-1} \sum_{m=1}^{M} F^{(m)}. \quad (19)$$

Clearly, these Bayesian estimates are consistent estimates of the corresponding posterior means; see Geyer (1992). It is rather difficult to derive analytic forms for the covariance matrices $\text{Var}(\alpha|X, Z)$, $\text{Var}(\theta|X, Z)$, and $\text{Var}(\xi|X, Z)$. However, estimates of these covariance matrices can be obtained as the corresponding sample covariance matrices based on the simulated observations. For example, a consistent estimate of $\text{Var}(\theta|X, Z)$ can be obtained as follows:

$$\hat{\text{Var}}(\theta|X, Z) = (M - 1)^{-1} \sum_{m=1}^{M} (\theta^{(m)} - \hat{\theta})(\theta^{(m)} - \hat{\theta})'.$$

Other statistical inferences, such as obtaining the confidence intervals and marginal distributions, can be carried out based on the simulated observations as well (see Besag et al., 1995; and Gilks, Richardson & Spiegelhalter, 1996).

4.2. Residuals and outlier statistics

The problem of residual and outlier analyses has been thoroughly studied and reviewed by Barnett & Lewis (1984), Hawkins (1980), and Cook & Weisberg (1982), among others. Chaloner & Brant (1988) developed some Bayesian outlier statistics for the linear model on the basis that the posterior distribution can be used to calculate the posterior probability that any observation is an outlier. For the normal linear model $y_i = x_i \beta + \epsilon_i$, where $y_i$ is the observed data, $x_i$ is the covariate, and $\epsilon_i$ is the residuals with distribution $N[0, \sigma^2]$; the $i$th
Assessing the plausibility of a proposed model is always fundamental in data analysis. The classical approach in structural equation modelling associated with non-Bayesian estimation, 4.3. Assessment of model

Again, observations with large \( |x_i| > K \sigma^2 \), for some choice of \( K \). The value of \( K \) can be selected so that the prior probability of an outlier is small and thus outliers are considered which are more extreme than is usually expected. For example, \( K \) may be chosen to be 2.0, 3.0 or \( \Phi^{-1}(0.5 + 0.5 \times 0.95^{1/m}) \). Hence the posterior outlier statistic \( \text{POS}(i; K) = P(|e_i|/\sigma > K|y_i) \) is used to provide a formal identification of outliers.

We now apply the approaches of Chaloner & Brant (1988), Albert & Chib (1993), Chaloner (1991), and Weiss (1994) to our nonlinear structural equation model. The essential idea is to study the residuals, and a particular observation whose residual is far from the expected value is regarded as an outlier. This is judged by the corresponding posterior outlier statistic. That is, the \( i \)th observation is regarded as an outlier if the corresponding posterior outlier statistic is large. This posterior outlier statistic can be directly estimated by using the observations \( \{(\alpha^{(m)}, \theta^{(m)}, F^{(m)}, Y^{(m)}), m = 1, \ldots, M\} \) generated by the hybrid algorithm. Informal identification of outliers can also be achieved by the plots of residuals (see Weiss & Lazaro, 1992). More details are given below.

Let \( e_{ki}^* = \psi_{ek}^{-1/2}(U_{ki} - \mu_k - \Lambda_{ek}^i \xi_i) \); from (1) and (2) \( e_{ki}^* \) can be regarded as the standardized residual. Let \( e_{ki}^{(m)*} = \psi_{ek}^{(m)-1/2}(U_{ki}^{(m)} - \mu_k^{(m)} - \Lambda_{ek}^{i(m)} \xi_i^{(m)}) \), where \( U_{i}^{(m)} = \{X, Y_i^{(m)}\} \), and \( \bar{e}_{ki} \) be the sample mean of \( \{e_{ki}^{(m)*}, m = 1, \ldots, M\} \). An informal identification of outliers can be obtained by plotting \( \bar{e}_{ki} \) against \( i \). The corresponding posterior outlier statistic is given by \( \text{POS}_{\varepsilon}(k, i; K) = P(|e_{ki}^*| > K|X, Z) \), for some selected value \( K \). An estimate of this statistic based on the generated observations is given by

\[
\hat{\text{POS}}_{\varepsilon}(k, i; K) = M^{-1} \sum_{m=1}^{M} I(|e_{ki}^{(m)*}| > K),
\]

(20)

where \( I(A) \) is the function which equals 1 if the statement \( A \) is true, and zero otherwise. This method requires the computation of \( np \) posterior outlier statistics in a data set. Computationally, it can be a moderate burden. An alternative way is to introduce multivariate outlier statistics. For example, it may be better to consider

\[
\text{POS}_{\varepsilon}(i; K) = P(e_{i}^{*} > K|X, Z),
\]

where \( e_{i}^{*} = \psi_{e}^{-1/2}(U_{i} - \mu - \Lambda_{e}^i \xi_i) \). Given the parameters, \( U_{i} \) and \( \xi_i \); \( e_{i}^{*} \) is distributed as \( \chi^2(p) \). So, it makes sense to choose \( K \) as \( \chi^2(p; 1 - \beta) \) for \( \beta = 0.01 \) or 0.05, where \( \chi^2(p, 1 - \beta) \) is the 100(1 - \( \beta \)) percentile of \( \chi^2(p) \). Let \( e_{i}^{(m)*} = \psi_{e}^{(m)-1/2}(U_{i}^{(m)} - \mu^{(m)} - \Lambda_{e}^{i(m)} \xi_i^{(m)}) \); \( \text{POS}_{\varepsilon}(i, k) \) can be similarly estimated by

\[
\hat{\text{POS}}_{\varepsilon}(i; K) = M^{-1} \sum_{m=1}^{M} I(e_{i}^{(m)*} > K).
\]

(21)

Again, observations with large \( \text{POS}_{\varepsilon} \) values can be identified as outliers. Finally, we can investigate the QQ plot (see Johnson & Wichern, 1992) of \( \varepsilon_{ki}^* \) to check the assumption of the normality. These analyses can also be similarly considered with variable \( \delta \) via the following statistic \( \text{POS}_{\delta}(i, k) = P(\delta_i^{-1} \delta_i > k|X, Z) \), where \( \delta_i = (I - \Pi)\xi_i^{(1)} - \Gamma \Theta (\xi_i^{(2)}) \).

4.3. Assessment of model

Assessing the plausibility of a proposed model is always fundamental in data analysis. The classical approach in structural equation modelling associated with non-Bayesian estimation,
such as maximum likelihood or generalized least squares, is to perform a goodness-of-fit test based on the asymptotic distribution of a test statistic that measures the discrepancy between
the posited model and the sample covariance matrix. When dealing with more complicated
nonlinear models, it is difficult to derive the distribution of such a statistic.

Meng (1994) introduced a Bayesian counterpart of the classical $p$-value by defining a
posterior predictive $p$-value that depends both on the data and the choice of priors. His
procedure is considered here to establish a goodness-of-fit assessment for the posited model
under the null hypothesis $H_0$ that the proposed model defined in (1) and (2) is plausible
(see also Zhu & Lee, 1999). More specifically, the posterior predictive $p$-value is
defined as

$$
p_B = \Pr(D(\mathbf{U}^{\text{rep}}|\theta, F, Y) \geq D(\mathbf{U}|\theta, F, Y)|X, Z, H_0),
$$
where $\mathbf{U}^{\text{rep}}$ denotes a replication of $\mathbf{U}$ and $D(\cdot|\cdot)$ is a discrepancy variable. The probability is
taken over the joint posterior distribution of $(\mathbf{U}^{\text{rep}}, \theta, F, Y)$ given $H_0$, $X$ and $Z$, where

$$
p(\mathbf{U}^{\text{rep}}, \theta, F, Y|X, Z, H_0) = p(\mathbf{U}^{\text{rep}}|\theta, F, Y)p(\theta, F, Y|X, Z).
$$

For our model, we choose the $\chi^2$ discrepancy variable

$$
D(\mathbf{U}^{\text{rep}}|\theta, F, Y) = \sum_{i=1}^{n} (U_i^{\text{rep}} - \mu - \Lambda \xi)^{\cdot} \Psi_{\epsilon}^{-1}(U_i^{\text{rep}} - \mu - \Lambda \xi),
$$
which is distributed as $\chi^2(pn)$. Here, implicitly, the partition $(\xi^{(1)}, \xi^{(2)})$ of $\xi$ is required to
satisfy the model as defined in (2). The posterior predictive $p$-value based on this discrepancy
variable is given by

$$
p_B(X, Z) = \int \{ \Pr(\chi^2(pn) \geq D(\mathbf{U}|\theta, F, Y)) | p(\theta, F, Y|X, Z) \} d\theta dF dY.
$$

A Rao–Blackwellized type estimate of $p_B(X, Z)$ is equal to

$$
\hat{p}_B(X, Z) = M^{-1} \sum_{m=1}^{M} \Pr(\chi^2(pn) \geq D(X, Y^{(m)}|\theta^{(m)}, F^{(m)})). \tag{22}
$$
The computation of $\hat{p}_B(X, Z)$ is straightforward, since $D(X, Y^{(m)}|\theta^{(m)}, Z^{(m)})$ can be calculated
in each iteration and the tail-area probability of the $\chi^2$ distribution can be obtained using any
standard statistical software. See Gelman, Meng & Stern (1996), and Meng (1994) for
more detailed discussions about the theoretical and practical aspects of the posterior
predictive $p$-value.

Another method to assess goodness of fit of the model is based on the following sum of
outlier indicator statistics:

$$
\text{SRE}_\epsilon(p_0) = \sum_{i=1}^{n} I\{\epsilon_i^* > \chi^2(p; 1 - p_0)\}.
$$
Conditional on $\xi$ and $U_i$, $\text{SRE}_\epsilon(p_0)$ is distributed as Binomial($n, p_0$) and has support on
$0, \ldots, n$. If the posterior distribution of $\text{SRE}_\epsilon(p_0)$ is on values that have large prior support,
then this statistic suggests the posited model fits the data. If the posterior lies partially on
implausible values, then there is some probability of lack of fit. If the posterior lies entirely on
implausible values, then lack of fit is identified. In practice, we apply this method via a normal
approximation of Binomial(n, p₀), and consider the posterior probability

\[ PP(p₀) = \Pr\{SRE(ε)(p₀) > np₀ + z₁₋β(np₀(1 - p₀))^{1/2}|X, Z, \} , \]

where β is chosen to be 0.01 or 0.05. An estimate of this posterior probability from the generated observations is given by

\[ \hat{PP}(p₀) = M^{-1} \sum_{m=1}^{M} I\{ SRE(ε)(p₀) > np₀ + z₁₋β(np₀(1 - p₀))^{1/2} \} , \]

with \( SRE(ε)(p₀) = \sum_{i=1}^{n} \{ ε_i^{(m)}ε_i^{(m)*} > χ^2(p; 1 - p₀) \} \). Large \( \hat{PP}(p₀) \) indicates that the model does not fit the data. Roughly, the proposed \( SRE(ε)(p₀) \) statistic plays a complementary role to the posterior predictive p-values.

5. Simulation study and example

5.1 Simulation Study

Results of a simulation study are presented here to give some idea of the empirical performance of the proposed Bayesian approach. A data set \( \{ U_i, i = 1, \ldots, n \} \) was generated from a nonlinear structural equation model defined in (1) and (2) with eight manifest variables which are related with four basic latent factors \( \xi^{(1)} = (ξ_{11}, ξ_{12}), \xi^{(2)} = (ξ_{i3}, ξ_{i4}), \) and \( H(ξ^{(2)}) = (ξ_{i3}, ξ_{i4}, ξ_{i5ξ_{i4}}, ξ_{i3ξ_{i3}}) \). Hence, some quadratic and interaction effects of the latent variables are considered. The continuous measurements \( U_{7i} \) and \( U_{8i} \) were transformed to polytomous observations via (3) with the thresholds \((-1.2^*, -0.6, 0.4, 1.0^*)\), where thresholds with an asterisk were held fixed in the estimation. For convenience, we used the same set of thresholds to define all the polytomous variables in the model. In this simulation study \( U_i = (X_i, Y_i) \), where the first six variables are continuous and the last two variables are polytomous. The matrices \( \mathbf{Γ}, \mathbf{Π} \) and \( \mathbf{Λ} \) are given by

\[
\mathbf{Γ} = \begin{bmatrix}
1.0 & -γ_{12} & γ_{13} & γ_{14} \\
γ_{21} & 1.0 & -γ_{23} & γ_{24}
\end{bmatrix}, \quad
\mathbf{Π} = \begin{bmatrix}
0 & 0 \\
π_{21} & 0
\end{bmatrix}, \quad
\mathbf{Λ}' = \begin{bmatrix}
1.0 & λ_{21} & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 1.0 & λ_{42} & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 1.0 & λ_{63} & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & λ_{84}
\end{bmatrix},
\]

where the ones and zeros are fixed known parameters, while \( π_{21}, λ_{ij} \) and \( γ_{ij} \) are unknown parameters. The true population values of the unknown parameters are given by: \( λ_{ij} = 0.8 \) and \( γ_{ij} = 0.6 \) for all \( i, j \) as specified in \( \mathbf{Λ} \) and \( \mathbf{Γ} \), \( π_{21} = 0.6 \), \( (Φ_{ξ11}, Φ_{ξ12}, Φ_{ξ22}) = (1.0, -0.5, 1.0) \), \( μ_j = 0.0 \), \( ψ_e = 1.0 \) for \( i = 1, \ldots, 8 \), \( ψ_{5k} = 1.0 \) for \( k = 1, 2 \). In this nonlinear model, there are 36 unknown parameters. Three sample sizes, \( n = 150, 300 \) and 600, were considered.

Three Bayesian estimates based on the following prior distributions were obtained:

BAY 1. Bayesian estimates based on conjugate priors with hyperparameters \( \{ μ_0, Λ_{0ek}, Γ_{0i} \} \) fixed at the true values; \( Σ_0 = I_8, ξ_{0ek} = α_{0ξi} = 10, β_{0ek} = β_{0ξi} = 8, H_{0ek} \) and \( H_{0k} \) are diagonal matrices with diagonal elements 0.25, \( ρ_0 = 8 \) and \( R_0^{-1} = 5Φ_{ξ} \). This can be regarded as a situation with good prior information.
BAY 2. Bayesian estimates based on conjugate priors with hyperparameters \( \{\mathbf{0}_{yk}, \mathbf{0}_{i} \} \) equal to 2.0 times the true values; \( \mathbf{0}_{yk} = 218, \mathbf{0}_{i} = I_8, \alpha_{0e} = \alpha_{0d} = 10, \beta_{0e} = \beta_{0d} = 8, \)
\( R_{0}(1, 1) = R_{0}(2, 2) = 1/5, R_{0}(1, 2) = 0.0, \mathbf{H}_{0e} \) and \( \mathbf{H}_{0d} \) are the same diagonal matrices as before. It can be seen that the prior values of \( \{\mathbf{0}, \mathbf{0}_{yk}, \mathbf{0}_{i} \} \) are quite different from those in BAY 1. This setting represents a situation with inaccurate prior information.

BAY 3. Bayesian estimates based on non-informative prior distributions.

The proposed hybrid algorithm with the Gibbs sampler and the MH algorithm is used to produce the Bayesian solutions in 100 replications. In the MH algorithm, we set \( \sigma = 1 \) and \( \sigma_{\alpha} = 0.1 \) for all \( k \) in the proposal distributions to give approximate acceptance rates 0.32 and 0.26, respectively. Based on three different starting values, parallel sequences of observations were generated and the EPSR values were computed. The algorithm is said to have converged if all the EPSR values are less than 1.2; see Cowles & Carlin (1996) and Gelman & Rubin (1992). We observed that in all cases, the algorithm converged within 2000 iterations. After the algorithm converged, a total of \( M = 2000 \) observations were collected with spacing \( s^* = 1 \). Then, the Bayesian estimates were computed via (19). The bias of each parameter estimate (which is the difference between the true value and the mean of the estimates based on 100 replications), the standard deviation of the estimates, and the root mean squares (RMS) between the estimates and the true value based on the replications are reported in Tables 1, 2 and 3. Here, the sums of the RMS across the estimates are presented in the last row. We have the following observations from these tables:

(i) BAY 1 estimates are better than BAY 2 estimates, but the differences are not significant.
Both BAY 1 and BAY 2 estimates are reasonably accurate; hence, it seems that the effect of the hyperparameter values is minor. Moreover, these estimates are better than BAY 3 estimates.

(ii) As expected, increasing the sample size improves the accuracy of the estimates and reduces the differences among the different types of Bayesian estimates.

(iii) Comparatively, estimates corresponding to the parameters relating to interaction and quadratic latent factors are less than accurate.

(iv) For \( n = 150 \), BAY 3 estimates associated with non-informative priors are not accurate.
Hence, it seems that conjugate priors are preferable for problems with small sample sizes.

5.2. An example from the World Values Survey

A small portion of the ICPSR data set collected by the World Values Survey 1981–1984 and 1990–1993 (World Values Study Group, 1994) is analysed in this example. The whole data set was collected in 45 societies around the world on broad topics such as work, religious belief, the meaning and purpose of life, family life, contemporary social issues, etc. As an illustration of our proposed method, only the data obtained from the United Kingdom were used. Six variables in the original data set (variables 180, 96, 62, 176, 116 and 117; see Appendix) that related to respondents’ employment, religious belief and homelife were taken as manifest variables in \( \mathbf{Y} = (y_1, \ldots, y_6)^T \). After deleting the cases with missing data, the sample size is 197. Among them, \( (y_1, y_2) \) are related to life, \( (y_3, y_4) \) are related to religious belief, and \( (y_5, y_6) \) are related to job satisfaction. Variables \( y_3 \) and \( y_4 \) are polytomous with five categories. Other variables were measured on a 10-point scale. Since we are using this example only to illustrate the methodology, these variables are treated as continuous variables.
A nonlinear structural equation model with the latent factors $\xi^{(1)} = (\xi_1)$ and $\xi^{(2)} = (\xi_2, \xi_3)$ is proposed with the following specifications: $\mathbf{H} = \mathbf{0}$, $\mathbf{H}(\xi^{(2)}) = (\xi_2, \xi_3, \xi_2\xi_3)$, and

$$
\mathbf{G}^* = \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \end{bmatrix}, \quad \mathbf{A}^* = \begin{bmatrix} 1.0 & \lambda_{21} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & \lambda_{42} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 & \lambda_{63} \end{bmatrix} ;
$$

### Table 1. Performance of the Bayesian estimates, $n = 150$

<table>
<thead>
<tr>
<th>Para.</th>
<th>BAY1</th>
<th>BAY2</th>
<th>BAY3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{1,2}$</td>
<td>-0.009</td>
<td>0.098</td>
<td>0.098</td>
</tr>
<tr>
<td>$\alpha_{1,3}$</td>
<td>0.009</td>
<td>0.091</td>
<td>0.092</td>
</tr>
<tr>
<td>$\alpha_{2,2}$</td>
<td>-0.009</td>
<td>0.105</td>
<td>0.105</td>
</tr>
<tr>
<td>$\alpha_{2,3}$</td>
<td>-0.018</td>
<td>0.098</td>
<td>0.100</td>
</tr>
<tr>
<td>$\pi_{21}$</td>
<td>-0.128</td>
<td>0.240</td>
<td>0.272</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>-0.059</td>
<td>0.205</td>
<td>0.213</td>
</tr>
<tr>
<td>$\gamma_{13}$</td>
<td>0.039</td>
<td>0.170</td>
<td>0.175</td>
</tr>
<tr>
<td>$\gamma_{14}$</td>
<td>-0.058</td>
<td>0.165</td>
<td>0.175</td>
</tr>
<tr>
<td>$\gamma_{21}$</td>
<td>0.010</td>
<td>0.155</td>
<td>0.155</td>
</tr>
<tr>
<td>$\gamma_{23}$</td>
<td>0.076</td>
<td>0.209</td>
<td>0.222</td>
</tr>
<tr>
<td>$\gamma_{24}$</td>
<td>0.010</td>
<td>0.197</td>
<td>0.198</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.010</td>
<td>0.064</td>
<td>0.065</td>
</tr>
<tr>
<td>$\lambda_{42}$</td>
<td>-0.002</td>
<td>0.061</td>
<td>0.061</td>
</tr>
<tr>
<td>$\lambda_{63}$</td>
<td>0.012</td>
<td>0.119</td>
<td>0.119</td>
</tr>
<tr>
<td>$\lambda_{84}$</td>
<td>0.019</td>
<td>0.149</td>
<td>0.150</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.021</td>
<td>0.194</td>
<td>0.195</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.006</td>
<td>0.161</td>
<td>0.161</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>-0.007</td>
<td>0.230</td>
<td>0.230</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>0.009</td>
<td>0.182</td>
<td>0.182</td>
</tr>
<tr>
<td>$\mu_5$</td>
<td>-0.023</td>
<td>0.116</td>
<td>0.119</td>
</tr>
<tr>
<td>$\mu_6$</td>
<td>-0.018</td>
<td>0.110</td>
<td>0.112</td>
</tr>
<tr>
<td>$\mu_7$</td>
<td>0.017</td>
<td>0.122</td>
<td>0.123</td>
</tr>
<tr>
<td>$\mu_8$</td>
<td>0.015</td>
<td>0.102</td>
<td>0.103</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>0.050</td>
<td>0.175</td>
<td>0.182</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>-0.026</td>
<td>0.134</td>
<td>0.136</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.004</td>
<td>0.201</td>
<td>0.201</td>
</tr>
<tr>
<td>$\psi_{61}$</td>
<td>0.013</td>
<td>0.186</td>
<td>0.186</td>
</tr>
<tr>
<td>$\psi_{62}$</td>
<td>-0.035</td>
<td>0.151</td>
<td>0.155</td>
</tr>
<tr>
<td>$\psi_{63}$</td>
<td>-0.017</td>
<td>0.166</td>
<td>0.167</td>
</tr>
<tr>
<td>$\psi_{64}$</td>
<td>-0.012</td>
<td>0.156</td>
<td>0.156</td>
</tr>
<tr>
<td>$\psi_{65}$</td>
<td>-0.016</td>
<td>0.139</td>
<td>0.140</td>
</tr>
<tr>
<td>$\psi_{66}$</td>
<td>-0.023</td>
<td>0.112</td>
<td>0.115</td>
</tr>
<tr>
<td>$\psi_{67}$</td>
<td>-0.030</td>
<td>0.168</td>
<td>0.170</td>
</tr>
<tr>
<td>$\psi_{68}$</td>
<td>0.013</td>
<td>0.181</td>
<td>0.181</td>
</tr>
<tr>
<td>$\Sigma \text{RMS}$</td>
<td>5.544</td>
<td>5.604</td>
<td>8.043</td>
</tr>
</tbody>
</table>
where the ones and zeros in $\mathbf{A}$ were treated as fixed parameters. To identify the parameters, we set $\alpha_{k,1} = \Phi^{-1}(n_{k,1}/n)$ and $\alpha_{k,4} = \Phi^{-1}(\Sigma_{i=1}^{4} n_{k,i}/n)$, where $n_{k,i}$ denotes the number of $y^k_i$. In this nonlinear model, there are a total of 26 structural parameters which are the elements in $\mu$ and $\Delta$, $\alpha_{k,i}$ ($k = 3, 4; i = 2, 3$), $\lambda_{\xi ij}$ in $\Lambda_{\xi}$, $\Phi_{\xi ij}(i \neq j)$ in $\Phi_{\xi}$, and the diagonal elements in $\Psi_{\xi}$ and $\Psi_{\delta}$. Bayesian estimates of these structural parameters and direct estimates of the basic latent factors were obtained via the proposed hybrid algorithm. The $\sigma$ and $\sigma_{\alpha_{i}}$ in the proposal distributions were set equal to 1.265 and 0.078, giving an approximately average
The following hyperparameters were selected: \( \alpha_{0\ell k} = \alpha_{0\ell b} = 10, \beta_{0\ell k} = \beta_{0\ell b} = 8, H_{0\ell k} \) and \( H_{0\ell k} \) are diagonal matrices with diagonal elements 0.25, \( \rho_0 = 20, \Sigma_0 = I_7, R_0^{-1} = 2\hat{\Phi}_{k}, \mu_0 = \hat{\mu} \) and \( \Lambda_{0k} = \hat{\Lambda}_k \), where \( \hat{\mu}, \hat{\Phi}_k \) and \( \hat{\Lambda}_k \) are the Bayesian estimates obtained using non-informative prior distributions. Based on different starting values of the parameters, three parallel sequences of observations were generated. Figure 1 presents the plots of the EPSR values against the iteration numbers. It can be seen that the hybrid algorithm converged after about 2000 iterations. To give more information about the
Figure 1. EPSR values of all parameters from three parallel runs in the ICPSR data.

Figure 2. Sample values of $\alpha_{3,2}$ and $\gamma_{13}$ from the hybrid algorithm for the ICPSR data.
convergence of the algorithm, plots of values of $\alpha_{3,2}$ and $\gamma_{13}$ against the iteration numbers are displayed in Figure 2. The convergent behaviours of other parameters are similar. After convergence, a total of $M = 2000$ observations with spacing $s^* = 1$ were collected for analysis.

Bayesian estimates of the structural parameters and their standard errors estimates are reported in Table 4. From the structure of $\Lambda$, latent factors $\xi_1$, $\xi_2$ and $\xi_3$ can be roughly interpreted as ‘life’, ‘religious belief’ and ‘job satisfaction’ factors; while $\xi^{(2)}\xi^{(3)}$ represents the interaction of ‘religious belief’ and ‘job satisfaction’. From $\gamma_{13}$ and its standard error, we see that the corresponding $t$-value is $-3.293$. It seems that the contribution of the interaction of ‘religious belief’ and ‘job satisfaction’ has a significant effect on ‘life’. Based on the proposed approach, other nonlinear terms such as $\xi^{(1)}\xi^{(1)}$, $\xi^{(2)}\xi^{(3)}$ etc., can also be analysed similarly by choosing appropriate structures for $\Lambda$ and $H(\xi)$. For completeness, direct Bayesian estimates of the basic factors $\xi_1$, $\xi_2$ and $\xi_3$ are displayed in Figure 3.

The posterior outlier statistics were calculated with $K = 3$. A few cases with large $\text{POS}_e(i; j, 3)$ and $\text{POS}_e(i; 16.81)$ values are summarized in Table 5, where $16.81 = \chi^2(6, 0.01)$. Observations with indices 7, 103 and 117 have large $\text{POS}_e(i; 16.81)$

---

**Table 4. The Bayesian estimates and their standard errors for the ICPSR data**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>EST</th>
<th>SD</th>
<th>Parameter</th>
<th>EST</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{3,2}$</td>
<td>0.305</td>
<td>0.057</td>
<td>$\mu_1$</td>
<td>8.414</td>
<td>0.121</td>
</tr>
<tr>
<td>$\alpha_{3,3}$</td>
<td>0.608</td>
<td>0.060</td>
<td>$\mu_2$</td>
<td>7.824</td>
<td>0.123</td>
</tr>
<tr>
<td>$\alpha_{4,2}$</td>
<td>-0.140</td>
<td>0.054</td>
<td>$\mu_3$</td>
<td>-0.001</td>
<td>0.081</td>
</tr>
<tr>
<td>$\alpha_{4,3}$</td>
<td>0.075</td>
<td>0.058</td>
<td>$\mu_4$</td>
<td>0.030</td>
<td>0.083</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.723</td>
<td>0.183</td>
<td>$\mu_5$</td>
<td>7.573</td>
<td>0.142</td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>0.710</td>
<td>0.142</td>
<td>$\mu_6$</td>
<td>7.430</td>
<td>0.167</td>
</tr>
<tr>
<td>$\gamma_{13}$</td>
<td>-0.494</td>
<td>0.150</td>
<td>$\psi_{e1}$</td>
<td>0.697</td>
<td>0.153</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.869</td>
<td>0.096</td>
<td>$\psi_{e2}$</td>
<td>1.287</td>
<td>0.191</td>
</tr>
<tr>
<td>$\lambda_{32}$</td>
<td>1.050</td>
<td>0.148</td>
<td>$\psi_{e3}$</td>
<td>0.447</td>
<td>0.085</td>
</tr>
<tr>
<td>$\lambda_{63}$</td>
<td>1.151</td>
<td>0.270</td>
<td>$\psi_{e4}$</td>
<td>0.528</td>
<td>0.092</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>0.519</td>
<td>0.122</td>
<td>$\psi_{e5}$</td>
<td>2.439</td>
<td>0.507</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>-0.110</td>
<td>0.108</td>
<td>$\psi_{e6}$</td>
<td>3.356</td>
<td>0.543</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>1.792</td>
<td>0.587</td>
<td>$\psi_{e1}$</td>
<td>0.828</td>
<td>0.168</td>
</tr>
</tbody>
</table>

---

Figure 3. Plots of the Bayesian estimates of the factor scores for the ICPSR data.
Table 5. The posterior outlier statistics for the ICPSR data

<table>
<thead>
<tr>
<th>Case number</th>
<th>$j = 1$</th>
<th>$j = 2$</th>
<th>$j = 3$</th>
<th>$j = 4$</th>
<th>$j = 5$</th>
<th>$j = 6$</th>
<th>$\text{POS}_{i}(i; 16.81)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0.205</td>
<td>0.556</td>
<td>0.008</td>
<td>0.006</td>
<td>0.485</td>
<td>0.151</td>
<td>1.000</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
<td>0.008</td>
<td>0</td>
<td>0.380</td>
<td>0.197</td>
</tr>
<tr>
<td>22</td>
<td>0.013</td>
<td>0.087</td>
<td>0</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0.025</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0.002</td>
<td>0.049</td>
<td>0.018</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
<td>0.003</td>
<td>0.001</td>
<td>0.387</td>
<td>0.126</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
<td>0</td>
<td>0.024</td>
<td>0</td>
<td>0.041</td>
</tr>
<tr>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
<td>0.017</td>
<td>0.158</td>
<td>0.048</td>
</tr>
<tr>
<td>65</td>
<td>0.173</td>
<td>0.023</td>
<td>0.004</td>
<td>0</td>
<td>0</td>
<td>0.006</td>
<td>0.455</td>
</tr>
<tr>
<td>74</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.092</td>
<td>0</td>
<td>0.053</td>
</tr>
<tr>
<td>77</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
<td>0.002</td>
<td>0.007</td>
<td>0.010</td>
<td>0.026</td>
</tr>
<tr>
<td>91</td>
<td>0</td>
<td>0</td>
<td>0.004</td>
<td>0.010</td>
<td>0</td>
<td>0.036</td>
<td>0.022</td>
</tr>
<tr>
<td>97</td>
<td>0</td>
<td>0.073</td>
<td>0.003</td>
<td>0</td>
<td>0.025</td>
<td>0</td>
<td>0.041</td>
</tr>
<tr>
<td>99</td>
<td>0.011</td>
<td>0</td>
<td>0.003</td>
<td>0</td>
<td>0.011</td>
<td>0</td>
<td>0.029</td>
</tr>
<tr>
<td>103</td>
<td>0.320</td>
<td>0.409</td>
<td>0.008</td>
<td>0.004</td>
<td>0</td>
<td>0.028</td>
<td>0.985</td>
</tr>
<tr>
<td>106</td>
<td>0.010</td>
<td>0</td>
<td>0.004</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
<td>0.023</td>
</tr>
<tr>
<td>117</td>
<td>0</td>
<td>0</td>
<td>0.002</td>
<td>0.009</td>
<td>0.020</td>
<td>0.505</td>
<td>0.547</td>
</tr>
<tr>
<td>120</td>
<td>0.232</td>
<td>0</td>
<td>0</td>
<td>0.012</td>
<td>0</td>
<td>0</td>
<td>0.166</td>
</tr>
<tr>
<td>123</td>
<td>0.057</td>
<td>0</td>
<td>0.002</td>
<td>0</td>
<td>0.003</td>
<td>0</td>
<td>0.072</td>
</tr>
<tr>
<td>140</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.014</td>
<td>0.111</td>
<td>0.032</td>
</tr>
<tr>
<td>149</td>
<td>0.013</td>
<td>0.046</td>
<td>0</td>
<td>0.005</td>
<td>0.019</td>
<td>0</td>
<td>0.165</td>
</tr>
<tr>
<td>154</td>
<td>0</td>
<td>0</td>
<td>0.011</td>
<td>0.017</td>
<td>0.047</td>
<td>0.005</td>
<td>0.259</td>
</tr>
<tr>
<td>158</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.053</td>
<td>0</td>
<td>0.035</td>
</tr>
<tr>
<td>169</td>
<td>0.013</td>
<td>0</td>
<td>0</td>
<td>0.004</td>
<td>0</td>
<td>0</td>
<td>0.036</td>
</tr>
<tr>
<td>192</td>
<td>0.006</td>
<td>0.066</td>
<td>0.004</td>
<td>0.004</td>
<td>0.095</td>
<td>0.003</td>
<td>0.421</td>
</tr>
</tbody>
</table>
values that are more than 0.50. The corresponding estimates of \( \text{POS}_e(i,j; 3) \) are also large, as expected. It may be concluded that the 7th, 103th and 117th observations are outliers. For completeness, plots of the posterior standardized residuals \( \tilde{e}_i^* \) against the observations indices are displayed in Figure 4. The statistic \( \text{SRE}_e(0.01) \) is calculated. For \( \beta = 0.01, 1.97 + z_{1-\beta} \sqrt{0.01 \times 0.99 \times 197} = 5.217 \). The posterior probability of \( \text{SRE}(0.01) > 5.217 \) is 0.510. The posterior predictive \( p \)-value (Gelman et al., 1996) for this example is equal to 0.503. These results indicate that the model fits the data well.

6. Concluding remarks

Models involving nonlinear effects are very common in social and behavioural sciences. However, in SEM, only a few empirical examples that incorporate nonlinear terms of latent variables into equations exist. As pointed out by Bollen & Paxton (1998), Schumacker & Marcoulides (1998), among others, the lack of applications is not due to the failure of substantive arguments that suggest the presence of nonlinearity, rather the existing statistical methods are technically demanding and not well understood. In our opinion, another reason may be the fact that rigorous statistical properties of most existing methods have not been established. In this paper, a Bayesian approach is proposed for analysing a general nonlinear LISREL-type model with mixed continuous and polytomous variables. In addition to point estimation, we provide statistical methods to obtain standard error estimates, residuals and outliers analyses, and model diagnoses using the posterior predictive check.

Owing to the complexity of the proposed model, the development of the Bayesian theory for analysis is non-trivial. As we have seen, difficulties arising from the nonlinear causal relationships among the latent variables and the discrete nature of the polytomous 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): 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. For example, as one reviewer kindly pointed out, the basic methodology developed here could possibly be applied to the following model, subject to the availability of identification conditions, with a nonlinear measurement model and a nonlinear structural model that is nonlinear in the coefficients as well:

\[
U_i = \mu + g_1(\Lambda, \xi_i) + \epsilon_i,
\]

\[
\xi_i^{(1)} = g_2(\Pi, \xi_i^{(1)}) + g_3(\Gamma, \xi_i^{(2)}) + \delta_i, \quad i = 1, \ldots, n,
\]

for given nonlinear functions \( g_1, g_2 \) and \( g_3 \). The basic idea is that once the augmented data associated with latent variables in \( \xi_i^{(1)} \) and \( \xi_i^{(2)} \) are given, the conditional model becomes the nonlinear regression model and hence the Gibbs sampler algorithm can be developed without much difficulty. Moreover, the strategy can also be applied to maximum likelihood estimation with EM-type algorithms; see, for example, Shi & Lee (1981, 1998), Lee & Tsang (1999), McCulloch (1997), and Zeger & Karim (1991), among many others. In the future, we expect to see further applications of this strategy to more complex psychometric models.
Acknowledgement

The authorship is alphabetical. This research has been supported by a Hong Kong UGC Earmarked grant, CUHK 4088/99H. The authors are indebted to the Editor and three anonymous reviewers for valuable comments in improving the paper, and also to ICPSR and the relevant funding agency for allowing the use of their data. The assistance of Michael K. H. Leung and Esther L. S. Tam is gratefully acknowledged.

References


Received 2 March 1999; revised version received 16 December 1999

**Appendix: Manifest variables in the ICPSR example**

The number of the variable corresponding to the original data set is given in parentheses at the end of each question.

- $y_1$: Overall, how satisfied are you with your home life? (V180)
- $y_2$: All things considered, how satisfied are you with your life as a whole these days? (V96)
- $y_3$: Religious beliefs? (V62)
- $y_4$: How important is God in your life? (V176)
- $y_5$: Overall, how satisfied or dissatisfied are you with your job? (V116)
- $y_6$: How free are you to make decisions in your job? (V117)