5 Some Practical Structural Equation Models

5.1 Introduction

In Chapter 2, we introduced standard linear and nonlinear SEMs and explained in details how to apply these models in practice. Bayesian methods for estimating parameters in the model and model comparison are presented in Chapters 3 and 4, respectively. While these developments provide sound statistical methods for solving many practical problems, they depend on certain assumptions. As these assumptions may not be satisfied by many complex data sets coming from substantive research, it is important to deal with those complex data structures.

In subsequent sections, we will introduce some generalizations of the standard SEMs for analyzing complex data sets. These include SEMs with mixed continuous and ordered categorical variables, SEMs with variables coming from an exponential family distribution, and SEMs with missing data. Moreover, we will illustrate how the Bayesian methodologies can be naturally extended to these generalizations, again through data augmentation and MCMC techniques.

5.2 SEMs with Continuous and Ordered Categorical Variables

5.2.1 Introduction

Due to the design of questionnaire and the nature of the problems on social, behavioral, and medical sciences, data often come from ordered categorical variables with observations in discrete form. Examples of such variables are attitude items, Likert items, rating scales and the like. A typical case is when a subject is asked to report the effect of a drug on scale like 'getting worse', 'no change', 'getting better'. One common approach in analyzing ordered categorical data is to treat the assigned integers as continuous data from a normal distribution. This approach may not lead to a serious problem if the histograms of the observations are symmetrical and with the highest frequencies at the center. This is, when most subjects choose the category 'no change'. To claim multivariate normality of the observed variables, we need to have most subjects choosing the middle category, in all the corresponding items. However, for interesting items in a questionnaire, most subjects would be likely to select categories at both ends. Hence, in practice, histograms corresponding to most variables are either skewed or bimodal; and routinely treating them as normal may lead to erroneous conclusions (see Olsson, 1979; Lee, Poon and Bentler, 1990).

A better approach for analyzing this kind of discrete data is to treat them as observations that come from a latent continuous normal distribution with a threshold specification. Suppose that for a given data set, the proportions of 0, 1, 2, 3 are 0.05, 0.05, 0.35, and 0.55, respectively. The discrete data are highly skewed to the left; see the histogram in Figure 5.1. The threshold approach for analyzing this highly skewed discrete variable is to treat the ordered categorical data as manifestations of an underlying normal variable y. The exact continuous measurements of y are not available, but are related to the observed ordered categorical variable z as follows: for $m = 0,\ 1,\ 2,\ 3,$

$$z = m \quad \text{if} \ \ \alpha_m \leq y < \alpha_{m+1};$$

where $-\infty = \alpha_0 < \alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 = \infty$, and $\alpha_1,\ \alpha_2$, and α_3 are thresholds. Then, the ordered categorical observations can be captured by $N[0, 1]$ with appropriate thresholds; see Figure 5.2. As $\alpha_2 - \alpha_1$ can be different from $\alpha_3 - \alpha_2$, unequal-interval scales are allowed. Thus, this threshold approach allows flexible modeling. As it is related to a common normal distribution, it also provides easy interpretation of the parameters.

Figures 5.1 and 5.2 here

Analysis of SEMs with mixed continuous and ordered categorical data is not straightforward, because we need to compute the multiple integrals associated with the cell probabilities that are induced by the ordered categorical outcomes. Some multistage methods have been proposed to reduce the computational burden in evaluating these integrals. The basic procedure of these multistage methods is: at the first stage partition the multivariate model to many bivariate submodels and estimate the polychoric and polyserial correlations, and the thresholds based on the bivariate submodels; then derive the asymptotic distributions of the estimates and analyze the SEM with a covariance structure analysis approach through a generalized least square (GLS) procedure. However, as pointed out by Shi and Lee (2000), the multistage estimators are not statistically optimal, and need to invert at each iteration of the GLS minimization a huge matrix whose dimension increases very rapidly with the number of observed variables.

The main objective of this section is to introduce a Bayesian approach for analyzing SEMs with mixed continuous and ordered categorical variables. The basic idea in handling the ordered categorical variables 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 data augmentation strategy, the model that is based on the complete data set becomes one with continuous variables. Again, in estimation, sequences of observations of the structural parameters, latent variables, and thresholds are simulated from the joint posterior distribution via MCMC algorithms. In addition to estimation, we also consider model selection via some Bayesian model selection statistics. Finally, an application related to quality of life (QOL) is presented to illustrate the Bayesian methodologies.

5.2.2 The Basic Model

Consider the following measurement equation for a $p \times 1$ observed random vector $\mathbf{v}_i$:

$$\mathbf{v}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \cdots, n, \tag{5.1}$$

where the definition of $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\omega}_i$ and $\boldsymbol{\epsilon}_i$ are the same as those in (2.3). Let $\boldsymbol{\eta}_i$ $(q_1 \times 1)$ and $\boldsymbol{\xi}_i$ $(q_2 \times 1)$ be latent subvectors of $\boldsymbol{\omega}_i$, and consider the following structural equation:

$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \tag{5.2}$$

where the definition of $\boldsymbol{\Pi}$ $(q_1 \times q_1)$, $\boldsymbol{\Gamma}$ $(q_1 \times q_2)$, $\boldsymbol{\xi}_i$, and $\boldsymbol{\delta}_i$ are the same as those in (2.8). Let $\boldsymbol{\Lambda}_\omega = (\boldsymbol{\Pi}, \boldsymbol{\Gamma})$, then Equation (5.2) can be written as $\boldsymbol{\eta}_i = \boldsymbol{\Lambda}_\omega \boldsymbol{\omega}_i + \boldsymbol{\delta}_i$. Let $\mathbf{v} = (\mathbf{x}^T, \mathbf{y}^T)^T$, where $\mathbf{x} = (x_1, \cdots, x_r)^T$ is a subset of variables whose exact continuous measurements are observable, while $\mathbf{y} = (y_1, \cdots, y_s)^T$ is the remaining subset of variables such that the corresponding continuous measurements are unobservable. The information associated with $\mathbf{y}$ is given by an observable ordered categorical vector $\mathbf{z} = (z_1, \cdots, z_s)^T$. Any latent variable in $\boldsymbol{\eta}$ or $\boldsymbol{\xi}$ may be related to observed variables in either $\mathbf{x}$ or $\mathbf{z}$. That is, any latent variable may have continuous and/or ordered categorical observed variables as its indicators. The relationship between $\mathbf{y}$ and $\mathbf{z}$ is defined by a set of thresholds as follows:

$$\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_s \end{bmatrix} \quad \text{if} \quad \begin{matrix} \alpha_{1,z_1} \le y_1 < \alpha_{1,z_1+1} \\ \vdots \\ \alpha_{s,z_s} \le y_s < \alpha_{s,z_s+1} \end{matrix}, \tag{5.3}$$

where for $k = 1, \cdots, s$, z_k is an integer value in $\{0, 1, \cdots, b_k\}$, and $\alpha_{k,0} < \alpha_{k,1} < \cdots < \alpha_{k,b_k} < \alpha_{k,b_k+1}$. In general, we set $\alpha_{k,0} = -\infty$, $\alpha_{k,b_k+1} = \infty$. For the kth variable, there are $b_k + 1$ categories which are defined by the unknown thresholds $\alpha_{k,j}$. The integer values $\{0, 1, \cdots, b_k\}$ of z_k are used for specifying the categories that contain the corresponding elements in y_k. In order to indicate the 'ordered' nature of the categorical values and their thresholds, it is better to choose an ordered set of integers for each z_k. In the Bayesian analysis, these integer values are neither directly used in the posterior simulation nor involved in any actual computation. In this section, the SEM with continuous and ordered categorical variables $\{(\mathbf{x}_i^T, \mathbf{z}_i^T)^T,\ i = 1, \cdots, n\}$ is analyzed.

The SEM defined by (5.1) and (5.2) is not identified without imposing appropriate identification conditions. There are two kinds of indeterminacies involved in this model. One is the common indeterminacy coming from the covariance structure of the model that can be solved with the common method of fixing appropriate elements in $\mathbf{\Lambda}, \mathbf{\Pi}$, and/or $\mathbf{\Gamma}$ at preassigned values. The other indeterminacy is induced by the ordered categorical variables. To tackle this identification problem, we should keep in mind the following two issues. First, obtaining a necessary and sufficient condition for identification is difficult, so we are mainly interested in finding a reasonable and convenient way to identify the model. Second, for an ordered categorical variable, the location and dispersion of its underlying continuous normal variable are unknown. As we have no idea about these latent values, it is desirable to take a unified scale to every ordered categorical variable, and to interpret the obtained statistical results in relative sense.

Now, let z_k be the ordered categorical variable that is defined with a set of thresholds and an underlying continuous variable y_k whose distribution is $N[\mu, \sigma^2]$. The indeterminacy is caused by the fact that the thresholds, μ and σ^2 are not simultaneously estimable. We follow a common practice (see for example, Shi and Lee, 1998) to impose identification conditions on the thresholds, which are the less interesting nuisance parameters. More specifically, we propose to fix the thresholds at both ends, $\alpha_{k,1}$ and α_{k,b_k}, at preassigned values. This method implicitly picks measures for the location and the dispersion of y_k. For instance, the range $\alpha_{k,b_k} - \alpha_{k,1}$ provides a standard for measuring the dispersion. This method can be applied to the multivariate case by imposing the above restrictions on the appropriate thresholds for every component in $\mathbf{z}$. If the model is scale invariant, the choice of the preassigned values for the fixed thresholds only changes the scale of the estimated covariance matrix (see Lee, Poon and Bentler, 1990). For better interpretation of the statistical results, it is advantageous to assign the values of the fixed thresholds so that the scale of each variable is the same. One common method is to use the observed frequencies and the standard normal distribution, $N[0,1]$. More specifically, for every k, we may fix $\alpha_{k,1} = \Phi^{*-1}(f^*_{k,1})$ and $\alpha_{k,b_k} = \Phi^{*-1}(f^*_{k,b_k})$, where $\Phi^*(\cdot)$ is the distribution function of $N[0,1]$, $f^*_{k,1}$ and f^*_{k,b_k} are the frequency of the first category, and the cumulative frequency of the category with $z_k < b_k$, respectively. For linear SEMs, these restrictions imply that the mean and the variance of the underlying continuous variable y_k are 0 and 1, respectively. For nonlinear SEMs, however, fixing $\alpha_{k,1}$ and α_{k,b_k} at preassigned values is only related to the location and dispersion of y_k, and the results should be interpreted with great caution.

5.2.3 Bayesian Analysis

We will utilize the useful strategy of data augmentation described in Chapter 3 in the Bayesian estimation for SEMs with continuous and ordered categorical variables. Let $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ and $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ be the observed continuous and ordered categorical data matrices, respectively; and let $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ and $\mathbf{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ be the matrices of latent continuous measurements and latent variables, respectively. The observed data $[\mathbf{X}, \mathbf{Z}]$ are augmented with the latent data $[\mathbf{Y}, \mathbf{\Omega}]$ in the posterior analysis. The joint Bayesian estimates of $\mathbf{\Omega}$, unknown thresholds in $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_s)$, and the structural parameter vector $\boldsymbol{\theta}$ that contains all unknown parameters in $\boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{\Lambda}, \mathbf{\Lambda}_\omega, \mathbf{\Psi}_\epsilon$, and $\mathbf{\Psi}_\delta$ will be obtained.

In the Bayesian estimation, we draw samples from the posterior distribution $[\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{\Omega}, \mathbf{Y}|\mathbf{X}, \mathbf{Z}]$ through the Gibbs sampler (Geman and Geman, 1984), which iteratively simulates $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$, $\mathbf{\Omega}$, and $\mathbf{Y}$ from the full conditional distributions. To implement the Gibbs sampler, we start with initial starting values $(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\theta}^{(0)}, \mathbf{\Omega}^{(0)}, \mathbf{Y}^{(0)})$, then simulate $(\boldsymbol{\alpha}^{(1)}, \boldsymbol{\theta}^{(1)}, \mathbf{\Omega}^{(1)}, \mathbf{Y}^{(1)})$ and so on according to the following procedure: at the jth iteration with current values $(\boldsymbol{\alpha}^{(j)},\ \boldsymbol{\theta}^{(j)},\ \mathbf{\Omega}^{(j)},\ \mathbf{Y}^{(j)})$,

$$\text{Step (a)}: \quad \text{Generate } \mathbf{\Omega}^{(j+1)} \text{ from } p(\mathbf{\Omega}|\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{X}, \mathbf{Z});$$

$$\text{Step (b)}: \quad \text{Generate } \boldsymbol{\theta}^{(j+1)} \text{ from } p(\boldsymbol{\theta}|\boldsymbol{\Omega}^{(j+1)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{X}, \mathbf{Z}); \tag{5.4}$$

$$\text{Step (c)}: \quad \text{Generate } (\boldsymbol{\alpha}^{(j+1)}, \mathbf{Y}^{(j+1)}) \text{ from } p(\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}^{(j+1)}, \boldsymbol{\Omega}^{(j+1)}, \mathbf{X}, \mathbf{Z}).$$

Convergence of the Gibbs sampler can be monitored by the 'estimated potential scale reduction (EPSR)' values suggested by Gelman (1996), or by plots of simulated sequences of the individual parameters with different starting points. Sequences of the quantities simulated from the joint posterior distribution will be used to calculate the Bayesian estimates and other related statistics. Based on the conjugate prior distributions of the parameters, conditional distributions required in the Gibbs sampler are presented in Appendix 5.1. The efficiency of the Gibbs sampler algorithm heavily depends on how easy one can sample observations from the conditional distributions. It can be seen from Appendix 5.1 that some of the conditional distributions are the familiar normal, Gamma, and inverted Wishart distributions. Drawing observations from these standard distributions is straightforward and fast. However, the joint conditional distribution $p(\boldsymbol{\alpha}, \mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\Omega})$ is nonstandard and complex. The Metropolis-Hastings (MH) algorithm (Metropolis *et al.*, 1953; Hastings, 1970) for sampling from $p(\boldsymbol{\alpha}, \mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\Omega})$ is also given in Appendix 5.1.

It has been shown (Geman and Geman, 1984; Geyer, 1992) that under mild conditions and for a sufficiently large j, the joint distribution of $(\boldsymbol{\alpha}^{(j)}, \boldsymbol{\theta}^{(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{Y}^{(j)})$ converges at an exponential rate to the desired posterior distribution $[\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Y}|\mathbf{X}, \mathbf{Z}]$. Hence, $[\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Y}|\mathbf{X}, \mathbf{Z}]$ can be approximated by the empirical distribution of a sufficiently large number of simulated observations collected after convergence of the algorithm. After obtaining a sufficiently large sample from the posterior distribution $[\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Y}|\mathbf{X}, \mathbf{Z}]$ with the MCMC algorithm, the Bayesian estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$, and $\boldsymbol{\Omega}$ can be obtained easily via the corresponding sample means of the simulated observations. Moreover, values of the goodness-of-fit and model comparison statistics can be obtained via the procedures as discussed in Chapter 4.

5.2.4 An Application: Bayesian Analysis of Quality of Life Data

There is increasing recognition that measures of quality of life (QOL) and health-related QOL have great value for clinical work and the planning and evaluation of health care as well as for medical research. It has been commonly accepted that QOL is a multidimensional concept (Staquet, Hayes and Fayes, 1998) that is best evaluated by a number of different latent constructs such as physical health status, mental health status, social relationships, and environmental conditions. As these latent constructs often cannot be measured objectively and directly, they are treated as latent variables in QOL analysis. The most popular method to assess a latent construct is using a survey which incorporates a number of related items that are intended to reflect the underlying latent construct of interest.

An exploratory factor analysis has been used as a method for exploring the structure of a new QOL instrument (The WHOQOL Group, 1998; Fayer and Machin, 1998), while a confirmatory factor analysis has been used to confirm the factor structure of the instrument. Recently, SEMs that are based on continuous observations with a normal distribution have been applied to QOL analysis (Power, Bullinger and Harper, 1999). Items in a QOL instrument are usually measured on an ordered categorical scale, typically in a 3- to 5-point scale. The discrete ordinal nature of the items also draws much attention in QOL analysis (Fayer and Machin, 1998; Fayer and Hand, 1997). It has been pointed out that non-rigorous treatment of ordinal items as continuous can be subjected to criticism (Glonek and McCullagh, 1995), and thus models such as the item response model and ordinal regression that incorporate the ordinal nature are more appropriate (Olschewski and Schumacker, 1990). The aim of this section is to apply the Bayesian methods for analyzing the following common QOL instrument with ordered categorical items.

The WHOQOL-BREF instrument is a shorten version of WHOQOL-100 (Power, Bullinger and Harper, 1999) by selecting 24 ordered categorical items out of the 100 items for evaluating four latent constructs. The first seven items (Q3 to Q9) are intended to address physical health, the next six items (Q10 to Q15) are intended to address psychological health, the three items (Q16, Q17, Q18) that follow measure social relationship, and the last eight items (Q19 to Q26) measure environment. In addition to the 24 ordered categorical items, this instrument also includes two ordered categorical items for the overall QOL (Q1) and overall health (Q2), giving a total of 26 items. All of the items are measured with a 5-point scale (1 = 'not at all/very dissatisfied'; 2 = 'a little/dissatisfied'; 3 = 'moderate/neither'; 4 = 'very much/satisfied'; 5 = 'extremely/very satisfied'). The sample size of the whole data set is extremely large. To illustrate the Bayesian methods, we only analyze a synthetic data set given in Lee (2007) with sample size $n = 338$. Appendix 1.1 presents the frequencies of all the ordered categorical items, which shows that many items are skewed to the left. Treating these ordered categorical data as coming from normal is problematic. Hence, the Bayesian approach that considers the discrete nature of the data is applied to the analysis of this ordered categorical data set.

Table 5.1 here

To illustrate the path sampling procedure, we compare an SEM with four explanatory latent variables to another SEM with three explanatory latent variables, see Lee *et al.* (2005). Let M_1 be the SEM whose measurement equation is defined by

$$\mathbf{y}_i = \mathbf{\Lambda}_1 \boldsymbol{\omega}_{1i} + \boldsymbol{\epsilon}_i, \tag{5.5}$$

where $\boldsymbol{\omega}_{1i} = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4})^T$, $\boldsymbol{\epsilon}_i$ is distributed as $N[\mathbf{0}, \boldsymbol{\Psi}_{\epsilon 1}]$, and

$$\boldsymbol{\Lambda}_1^T = \begin{bmatrix} 1 & \lambda_{21} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \lambda_{42} & \cdots & \lambda_{92} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_{11,3} & \cdots & \lambda_{15,3} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_{17,4} & \lambda_{18,4} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 1 & \lambda_{20,5} & \cdots & \lambda_{26,5} \end{bmatrix}.$$

The structural equation of M_1 is defined by

$$\eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \gamma_4\xi_{i4} + \delta_i, \tag{5.6}$$

where $(\xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4})^T$ and δ_i are independently distributed as $N[\mathbf{0}, \boldsymbol{\Phi}_1]$ and $N[0, \sigma^2_{\delta 1}]$, respectively. Let M_2 be the SEM whose measurement equation is defined by

$$\mathbf{y}_i = \boldsymbol{\Lambda}_2\boldsymbol{\omega}_{2i} + \boldsymbol{\epsilon}_i, \tag{5.7}$$

where $\boldsymbol{\omega}_{2i} = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3})^T$, and $\boldsymbol{\epsilon}_i$ is distributed according to $N[\mathbf{0}, \boldsymbol{\Psi}_{\epsilon 2}]$. The first three columns of $\boldsymbol{\Lambda}_2$ are the same as those given in $\boldsymbol{\Lambda}_1$ except without the rows corresponding to Q19 to Q26, while the last column is given by $[0, \cdots, 0, 1, \lambda_{17,4}, \lambda_{18,4}]^T$. The structural equation of M_2 is defined by

$$\eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \delta_i, \tag{5.8}$$

where $(\xi_{i1}, \xi_{i2}, \xi_{i3})^T$ and δ_i are independently distributed as $N[\mathbf{0}, \boldsymbol{\Phi}_2]$ and $N[0, \sigma^2_{\delta 2}]$, respectively. The Bayesian analysis is conducted using the conjugate prior distributions. The hyperparameter values corresponding to the prior distributions of the unknown loadings in $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are all taken to be 0.8; those corresponding to $\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ are $\{0.6, 0.6, 0.4, 0.4\}$; those corresponding to $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ are $\rho_0 = 30$ and $\mathbf{R}_0^{-1} = 8\mathbf{I}$; other hyperparameter values are taken as $\mathbf{H}_{0yk} = 0.25\mathbf{I}, \mathbf{H}_{0\omega k} = 0.25\mathbf{I}$, $\alpha_{0\epsilon k} = \alpha_{0\delta k} = 10$, and $\beta_{0\epsilon k} = \beta_{0\delta k} = 8$, where $\mathbf{I}$ is a generic notation indicating the identity matrix with an appropriate dimension. In the path sampling procedure in computing the Bayes factor, we take $S = 10$, and $J = 2,000$ after a 'burn-in' phase of 1,000 iterations.

It is not easy to find a link model M_t that directly links M_1 and M_2. Hence, we first compare M_1 with the following simple model M_0:

$$M_0 : \quad \mathbf{y}_i = \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{\epsilon}_i \overset{D}{=} N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$, and $\boldsymbol{\Psi}_\epsilon$ is a diagonal matrix. The measurement equation of the link model of M_0 and M_1 is defined by $M_t : \ \mathbf{y}_i = t\boldsymbol{\Lambda}_1\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$, and the structural equation is given by (5.6). We obtain $\widehat{\log B_{10}} = 81.05$. Similarly, we compare M_2 and M_0 via the path sampling procedure, and find that $\widehat{\log B_{20}} = 57.65$. From the above results, we can get an estimate of $\widehat{\log B_{12}} = \widehat{\log B_{10}} - \widehat{\log B_{20}}$, which is equal to 23.40. Hence, M_1, the SEM with four explanatory latent variables, is selected. Bayesian estimates of the unknown structural parameters in M_1 are presented in Figure 5.3. The less interesting threshold estimates are not presented. All the factor loading estimates, except $\hat{\lambda}_{17,4}$ that associates with the indicator 'sexual activity', are high. This indicates a strong association between each of the latent variables and their corresponding indicators. From the meaning of the items, η, ξ_1, ξ_2, ξ_3 and ξ_4 can be interpreted as the overall QOL, physical health, psychological health, social relationship, and environment, respectively. The estimates of correlations $\{\phi_{12}, \phi_{13}, \phi_{14}, \phi_{23}, \phi_{24}, \phi_{34}\}$ among the explanatory latent variables are equal to $\{0.46, 0.29, 0.42, 0.48, 0.51, 0.48\}$, indicating that these explanatory latent variables are highly correlated. The estimated structural equation that addresses the relations of QOL with the latent constructs about physical and psychological health, social relationship, and environment is

$$\eta = 0.72\xi_1 + 0.32\xi_2 + 0.17\xi_3 - 0.04\xi_4.$$

Thus, physical health has the most important effect on QOL, followed by psychological health and social relationship, while the effect of environment is not important.

Figure 5.3 here

The software WinBUGS (Spiegelhalter *et al.*, 2003) can produce Bayesian estimates of the structural parameters and latent variables for SEMs with ordered categorical variables. According to our understanding of WinBUGS, it is not straightforward to apply this software to simultaneously estimate the unknown thresholds and structural parameters. Hence, in applying WinBUGS, we first estimate all the thresholds through the method as described in Section 5.2.2. Then, the thresholds are fixed in the WinBUGS program in producing the Bayesian solutions. Note that this procedure may underestimate the standard errors. Hence, hypothesis testing should be conducted through DIC, rather than the z-score that depends on the standard error estimate. The WinBUGS code and the data related to the above QOL analysis are respectively given in the following websites:

http://www.sta.cuhk.edu.hk/song-lee/book-chapter5(section5.2.4)/WinBUGS-code,

http://www.sta.cuhk.edu.hk/song-lee/book-chapter5(section5.2.4)/WinBUGS-data.

(PLEASE CHANGE TO A WEB-SITE HOUSED IN JOHN-WILEY).

5.2.5 SEMs with Dichotomous Variables

In this subsection, we will focus on dichotomous variables that are ordered binary and defined with one threshold. Dichotomous variables arise when respondents are asked to select answers from 'Yes or No' about the presence of a symptom, 'Feeling better or Worse' about the effect of a drug, 'True or False' about a test item, etc. The usual numerical values assigned to these variables are the ad hoc numbers with an ordering such as '0' and '1', or '1' and '2'. In analyzing dichotomous data, the basic assumption in SEM that the data come from a continuous normal distribution is clearly violated, and rigorous analysis that takes into account the dichotomous nature is necessary. Analysis of SEMs with dichotomous variables is similar to but not exactly the same as the analysis with ordered categorical variables.

In many substantive researches, particularly in education, it is important to explore and determine a small number of intrinsic latent factors under a number of test items. Item factor analysis is an important model that has been proposed for explaining the underlying factor structures; see Bock and Aitkin (1981), Meng and Schilling (1996), and Lee and Song (2003). Another direction of analysis is motivated by the fact that correlated dichotomous data arise frequently in many medical and biological studies, ranging from measurements of random cross-section subjects to repeated measurements in longitudinal studies. A popular model is the multivariate probit (MP) model in biostatistics and other fields. This model is described in terms of a correlated multivariate normal distribution of the underlying latent variables that are manifested as discrete variables, again through a threshold specification. The emphasis of the MP model is on the mean structure, and the main difficulty in the analysis is on evaluating the multivariate normal orthant probabilities induced by the dichotomous variables; see Gibbons and Wilcox-Gök (1998), and Chib and Greenberg (1998). Analysis of the MP model requires the simulation of observations from a multivariate truncated normal distribution with an arbitrary covariance matrix. Even with the efficient methods in statistical computing, the underlying computational effort is heavy. We will show below that this computational burden can be reduced by adopting an SEM approach.

Consider a common SEM defined by (5.1) and (5.2). Now suppose that the exact measurement of $\mathbf{y}_i = (y_{i1}, \cdots, y_{ip})^T$ is not available and its information is given by an observed dichotomous vector $\mathbf{z}_i = (z_{i1}, \cdots, z_{ip})^T$ such that for $k = 1, \cdots, p$,

$$z_{ik} = \begin{cases} 1, & \text{if } \ y_{ik} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{5.9}$$

The available observed data set is $\{\mathbf{z}_i, \ i = 1, \cdots, n\}$. Consider the relationship between the measurement equation model defined by (5.1) and the dichotomous variables in $\mathbf{z}$.

Let $\mathbf{\Lambda}_k^T$, μ_k, and $\psi_{\epsilon k}$ be the kth row of $\mathbf{\Lambda}$, the kth element of $\boldsymbol{\mu}$, and the kth diagonal element of $\mathbf{\Psi}_\epsilon$, respectively. It follows from (5.9) that

$$\Pr(z_{ik} = 1|\boldsymbol{\omega}_i, \mu_k, \mathbf{\Lambda}_k, \psi_{\epsilon k}) = \Pr(y_{ik} > 0|\boldsymbol{\omega}_i, \mu_k, \mathbf{\Lambda}_k, \psi_{\epsilon k}) = \Phi^*\{(\mathbf{\Lambda}_k^T/\psi_{\epsilon k}^{1/2})\boldsymbol{\omega}_i \ + \mu_k/\psi_{\epsilon k}^{1/2}\},$$

where $\Phi^*(\cdot)$ is the distribution function of $N[0,1]$. Note that $\mu_k, \mathbf{\Lambda}_k$, and $\psi_{\epsilon k}$ are not estimable, because $C\mathbf{\Lambda}_k^T/(C\psi_{\epsilon k}^{1/2}) = \mathbf{\Lambda}_k^T/\psi_{\epsilon k}^{1/2}$ and $C\mu_k/(C\psi_{\epsilon k}^{1/2}) = \mu_k/\psi_{\epsilon k}^{1/2}$ for any positive constant C. There are many ways to solve this identification problem. Here, we fix $\psi_{\epsilon k} = 1.0$. Note that the value 1.0 is chosen for convenience, and any other value would give an equivalent solution up to a change of scale. Again the measurement and structural equations are identified by fixing the approximate elements of $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\omega$ at preassigned values.

Let $u_{ik} = \mathbf{\Lambda}_k^T\boldsymbol{\omega}_i + \epsilon_{ik}$, that is $u_{ik} + \mu_k = y_{ik}$. As $y_{ik} \geq 0$ if and only if $u_{ik} \geq -\mu_k$, or $z_{ik} = 1$ if and only if $u_{ik} \geq -\mu_k$. Consequently, $-\mu_k$ is the threshold corresponding to u_{ik}. Because there are at least two thresholds associated with an ordered categorical variable, the relation of the thresholds and μ_k is not as clear. Also, the identification conditions are slightly different. Thus, methods for analyzing these two types of discrete variables are not exactly the same.

Let $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ be the observed data set of dichotomous variables, let $\boldsymbol{\theta}$ be the unknown parameter vector, which contains parameters in $\boldsymbol{\mu}$, $\mathbf{\Lambda}$, $\mathbf{\Lambda}_\omega$, $\mathbf{\Phi}$, and $\mathbf{\Psi}_\delta$, let $\mathbf{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ be the matrix of latent variables in the model, and $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ be the matrix of latent continuous measurements underlying the matrix of observed dichotomous data $\mathbf{Z}$. In the Bayesian analysis, the observed data $\mathbf{Z}$ is augmented with $\mathbf{\Omega}$ and $\mathbf{Y}$; and a large sample of observations will be sampled from $p(\boldsymbol{\theta}, \mathbf{\Omega}, \mathbf{Y}|\mathbf{Z})$ through the following Gibbs sampler (Geman and Geman, 1984). At the jth iteration with current values $\boldsymbol{\theta}^{(j)}$, $\mathbf{\Omega}^{(j)}$, and $\mathbf{Y}^{(j)}$:

$$\text{Step (a)}: \quad \text{Generate } \mathbf{\Omega}^{(j+1)} \text{ from } p(\mathbf{\Omega}|\boldsymbol{\theta}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{Z}),$$

$$\text{Step (b)}: \quad \text{Generate } \boldsymbol{\theta}^{(j+1)} \text{ from } p(\boldsymbol{\theta}|\boldsymbol{\Omega}^{(j+1)}, \mathbf{Y}^{(j)}, \mathbf{Z}),$$

$$\text{Step (c)}: \quad \text{Generate } \mathbf{Y}^{(j+1)} \text{ from } p(\mathbf{Y}|\boldsymbol{\theta}^{(j+1)}, \boldsymbol{\Omega}^{(j+1)}, \mathbf{Z}).$$

Conditional distributions related to Step (a) and (b) can be obtained similarly as before. For $p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Z})$, as the $\mathbf{y}_i$ are mutually independent, it follows that

$$p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Z}) = \prod_{i=1}^{n} p(\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\omega}_i, \mathbf{z}_i).$$

Moreover, it follows from the definition of the model and (5.9) that

$$[y_{ik}|\boldsymbol{\theta}, \boldsymbol{\omega}_i, \mathbf{z}_i] \stackrel{D}{=} \begin{cases} N[\mu_k + \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i, 1] I_{(-\infty,0]}(y_{ik}), & \text{if } z_{ik} = 0, \\ N[\mu_k + \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i, 1] I_{(0,\infty)}(y_{ik}), & \text{if } z_{ik} = 1, \end{cases} \tag{5.10}$$

where $I_A(y)$ is an indicator function that takes the value 1 if y is in A, and 0 otherwise.

Statistical inference of the model can be obtained on the basis of the simulated sample of observations from $p(\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Y}|\mathbf{Z})$ as before. Note that for dichotomous data analysis, a vector of dichotomous observation $\mathbf{z}_i$ rather than $\mathbf{y}_i$ is observed; hence a lot of information of $\mathbf{y}$ is lost. As a result, it requires a large sample size to achieve accurate estimates. The PP p-value (Gelman, Meng and Stern, 1996) for assessing the goodness-of-fit of a posited model, and the Bayes factor for model comparison can be obtained via similar developments as before.

5.3 SEMs with Variables from Exponential Family Distributions

5.3.1 Introduction

The standard SEMs discussed so far are developed under a crucial assumption that the conditional distribution of observed variables given latent variables is normal. For example, although we do not assume that $\mathbf{y}$ is normal in a nonlinear SEM, we assume that $\mathbf{y}|\boldsymbol{\omega}$ or $\boldsymbol{\epsilon}$ is normal. In this section, we generalize the distribution of $\mathbf{y}|\boldsymbol{\omega}$ from normal to the exponential family distributions (EFDs). This family is very general, it includes discrete distributions such as binomial and Poisson, and continuous distributions such as normal and Gamma, as special cases. The EFDs have been extensively used in many areas of statistics; particularly in relation to latent variable models, such as the generalized linear models (McCullagh and Nelder, 1989), and generalized linear mixed models (Booth and Hobert, 1999). In contrast to SEMs, the main objective of these latent variable models is to assess the effects of covariates on the observed variables, and the latent variables in these models are usually used to model the random effects. Motivated by the above consideration, in this section, we will consider a nonlinear SEM that can accommodate covariates, variables from the EFDs, and ordered categorical variables. The strategy that combines the data augmentation and MCMC methods is again used to develop the Bayesian methodologies.

5.3.2 The SEM Framework with EFDs

A nonlinear SEM with fixed covariates on the basis of EFDs is defined as follows (see Song and Lee, 2007). For $i = 1, \cdots, n$, let $\mathbf{y}_i = (y_{i1}, \cdots, y_{ip})^T$ be a vector of observed variables measured on each of the n independently distributed individuals. For brevity, we assume that the dimension of $\mathbf{y}_i$ is the same for each i, however, this assumption can be relaxed without much difficulty. We wish to identify the relationship between the observed variables in $\mathbf{y}_i$ and the related latent vector $\boldsymbol{\omega}_i = (\omega_{i1}, \cdots, \omega_{iq})^T$ with fixed covariates. For $k = 1, \cdots, p$, we assume that y_{ik} given $\boldsymbol{\omega}_i$ are independent and the corresponding conditional distributions come from the following exponential family with a canonical parameter ϑ_{ik} (Sammel, Ryan and Legler, 1997):

$$\begin{aligned} p(y_{ik}|\boldsymbol{\omega}_i) &= \exp\{[y_{ik}\vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\epsilon k} + c_k(y_{ik}, \psi_{\epsilon k})\}, \\ E(y_{ik}|\boldsymbol{\omega}_i) &= \dot{b}(\vartheta_{ik}), \quad \text{and} \quad \mathrm{Var}(y_{ik}|\boldsymbol{\omega}_i) = \psi_{\epsilon k}\ddot{b}(\vartheta_{ik}), \end{aligned} \tag{5.11}$$

where $b(\cdot)$ and $c_k(\cdot)$ are specific differentiable functions with the dots denoting the derivatives, and $\vartheta_{ik} = g_k(\mu_{ik})$ with a link function g_k. Let $\boldsymbol{\vartheta}_i = (\vartheta_{i1}, \cdots, \vartheta_{ip})^T$, $\mathbf{c}_{ik}$ $(m_k \times 1)$

be vectors of fixed covariates, $\mathbf{A}_k$ be a $m_k \times 1$ vector of unknown parameters, and $\mathbf{\Lambda} = (\mathbf{\Lambda}_1, \cdots, \mathbf{\Lambda}_p)^T$ is a matrix of unknown parameters. We use the following model to analyze the relation of ϑ_{ik} with $\mathbf{c}_{ik}$ and $\boldsymbol{\omega}_i$. For $k = 1, \cdots, p$,

$$\vartheta_{ik} = \mathbf{A}_k^T \mathbf{c}_{ik} + \mathbf{\Lambda}_k^T \boldsymbol{\omega}_i. \tag{5.12}$$

Note that Equation (5.12) can accommodate an intercept μ_k by taking a component of $\mathbf{c}_{ik}$ as 1 and defining the corresponding component of $\mathbf{A}_k$ as μ_k. This equation can be viewed as a 'measurement' model. Its main purpose is to identify the latent variables via the corresponding observed variables in $\mathbf{y}$, with the help of the fixed covariates $\mathbf{c}_{ik}$. See Section 5.3.4 for a concrete example. Let $\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$, where $\boldsymbol{\eta}_i$ is the outcome latent vector, and $\boldsymbol{\xi}_i$ is the explanatory latent vector. To assess how the latent variables affect each other, we introduce the following nonlinear structural equation with fixed covariates:

$$\boldsymbol{\eta}_i = \mathbf{B}\mathbf{d}_i + \mathbf{\Pi}\boldsymbol{\eta}_i + \mathbf{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i, \tag{5.13}$$

where $\mathbf{d}_i$ is a vector of fixed covariates, $\mathbf{F}(\boldsymbol{\xi}_i) = (f_1(\boldsymbol{\xi}_i), \cdots, f_l(\boldsymbol{\xi}_i))^T$, $f_j(\boldsymbol{\xi}_i)$ is a nonzero differentiable function of $\boldsymbol{\xi}_i$, $\boldsymbol{\delta}_i$ is an error term, and $\mathbf{B}, \mathbf{\Pi}$, and $\mathbf{\Gamma}$ are unknown parameter matrices. Fixed covariates in $\mathbf{d}_i$ may or may not be equal to those in $\mathbf{c}_{ik}$. The distributions of $\boldsymbol{\xi}_i$ and $\boldsymbol{\delta}_i$ are $N[\mathbf{0}, \mathbf{\Phi}]$ and $N[\mathbf{0}, \mathbf{\Psi}_\delta]$, respectively, and $\boldsymbol{\xi}_i$ and $\boldsymbol{\delta}_i$ are uncorrelated. For computing efficiency and stability, the covariance matrix $\mathbf{\Psi}_\delta$ is assumed to be diagonal. We assume that $\mathbf{I} - \mathbf{\Pi}$ is nonsingular and its determinant is independent of elements in $\mathbf{\Pi}$. Let $\mathbf{\Lambda}_\omega = (\mathbf{B}, \mathbf{\Pi}, \mathbf{\Gamma})$, and $\mathbf{G}(\boldsymbol{\omega}_i) = (\mathbf{d}_i^T, \boldsymbol{\eta}_i^T, \mathbf{F}(\boldsymbol{\xi}_i)^T)^T$, then (5.13) can be rewritten as $\boldsymbol{\eta}_i = \mathbf{\Lambda}_\omega \mathbf{G}(\boldsymbol{\omega}_i) + \boldsymbol{\delta}_i$.

To accommodate ordered categorical data, we allow any component y of $\mathbf{y}$ to be unobservable, and its information is given by an observable ordered categorical variable z as follows: $z = m$ if $\alpha_m \leq y < \alpha_{m+1}$, for $m = 0, \cdots, b$, where $\{-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_b < \alpha_{b+1} = \infty\}$ is the set of thresholds that defines the categories. This model can be identified as before. For instance, each ordered categorical variable can be identified by fixing α_1 and α_b at preassigned values; and the SEM can be identified by the common practice of restricting the appropriate elements in $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\omega$ to fixed known values. Dichotomous variables can be analyzed as a special case of ordered categorical variables with slight modifications.

5.3.3 Bayesian Inference

Again, Bayesian methods are developed via the useful strategy that combines data augmentation and MCMC methods. In this subsection, we present the full conditional distributions in the implementation of the Gibbs sampler for simulating observations of the parameters and the latent variables from their joint posterior distribution. These simulated observations are used to obtain the Bayesian estimates and their standard error estimates, and to compute the Bayes factor for model comparison.

Proper conjugate prior distributions are taken for various unknown parameters. More specifically, let $\mathbf{\Psi}_\epsilon$ be the diagonal covariance matrix of the error measurements that correspond to the ordered categorical variables:

$$\begin{aligned}&\mathbf{A}_k \overset{D}{=} N[\mathbf{A}_{0k}, \mathbf{H}_{0k}],\ \psi_{\epsilon k}^{-1} \overset{D}{=} Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}],\ [\mathbf{\Lambda}_k|\psi_{\epsilon k}] \overset{D}{=} N[\mathbf{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}],\\ &\mathbf{\Phi}^{-1} \overset{D}{=} W_{q_2}[\mathbf{R}_0, \rho_0],\ \psi_{\delta k}^{-1} \overset{D}{=} Gamma[\alpha_{0\delta k}, \beta_{0\delta k}],\ [\mathbf{\Lambda}_{\omega k}|\psi_{\delta k}] \overset{D}{=} N[\mathbf{\Lambda}_{0\omega k}, \psi_{\delta k}\mathbf{H}_{0\omega k}].\end{aligned} \tag{5.14}$$

where $\psi_{\epsilon k}$ and $\psi_{\delta k}$ are the kth diagonal element of $\mathbf{\Psi}_\epsilon$ and $\mathbf{\Psi}_\delta$, respectively; $\mathbf{\Lambda}_k^T$ and $\mathbf{\Lambda}_{\omega k}^T$ are the kth rows of $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\omega$, respectively; $\mathbf{A}_{0k}, \alpha_{0\epsilon k}, \beta_{0\epsilon k}, \mathbf{\Lambda}_{0k},\ \alpha_{0\delta k},\ \beta_{0\delta k}, \mathbf{\Lambda}_{0\omega k},\ \rho_0$, and positive definite matrices $\mathbf{H}_{0k}, \mathbf{H}_{0yk}, \mathbf{H}_{0\omega k}$, and $\mathbf{R}_0$ are hyperparameters whose values are assumed to be given by the prior information. For $k \neq l$, it is assumed that prior distributions of $(\psi_{\epsilon k}, \mathbf{\Lambda}_k)$ and $(\psi_{\epsilon l}, \mathbf{\Lambda}_l)$, $(\psi_{\delta k}, \mathbf{\Lambda}_{\omega k})$ and $(\psi_{\delta l}, \mathbf{\Lambda}_{\omega l})$, as well as $\mathbf{A}_k$ and $\mathbf{A}_l$ are independent.

We first consider the situation in which components in $\mathbf{y}_i$ are neither dichotomous nor ordered categorical, but can be directly observed. Let $\mathbf{Y}$ be the observed data set. Based on the idea of data augmentation, we focus on the joint posterior distribution $[\mathbf{\Omega}, \boldsymbol{\theta}|\mathbf{Y}]$, where $\mathbf{\Omega}$ contains all the latent vectors, and $\boldsymbol{\theta}$ is the parameter vector that contains all the unknown parameters in the model. The Gibbs sampler (Geman and Geman, 1984) is used to simulate observations from the posterior distribution $[\mathbf{\Omega}, \boldsymbol{\theta}|\mathbf{Y}]$. The required full conditional distributions are given in Appendix 5.2.

To handle the ordered categorical data, let $\mathbf{y}_k^{*^T}$ be the kth row of $\mathbf{Y}$ that is not directly observable. Let $\mathbf{z}_k$ be the corresponding observable ordered categorical vector, and $\boldsymbol{\alpha}_k = (\alpha_{k,2}, \cdots, \alpha_{k,b_k-1})$. It is natural to assume that the prior distribution of $\boldsymbol{\alpha}_k$ is independent of the prior distribution of $\boldsymbol{\theta}$. To deal with a general situation in which there is little or no information about the thresholds, the following noninformative prior distribution is used: $p(\boldsymbol{\alpha}_k) = p(\alpha_{k,2}, \cdots, \alpha_{k,b_k-1}) \propto C$ for $\alpha_{k,2} < \cdots < \alpha_{k,b_k-1}$, where C is a constant. Moreover, it is assumed that $\boldsymbol{\alpha}_k$ and $\boldsymbol{\alpha}_l$ are independent for $k \neq l$. Based on these assumptions, the full conditional distribution $[\boldsymbol{\alpha}_k, \mathbf{y}_k^*|\mathbf{z}_k, \mathbf{\Omega}, \boldsymbol{\theta}]$ is derived and presented in Appendix 5.2.

The Gibbs sampler algorithm proceeds by sampling $\boldsymbol{\omega}_i$, $(\boldsymbol{\alpha}_k, \mathbf{y}_k^*)$, and $\boldsymbol{\theta}$ from their full conditional distributions, most of which are nonstandard because of the assumption of EFDs. Various forms of the Metropolis-Hastings (MH) algorithm (Metropolis *et al.*, 1953; Hastings, 1970) are required to simulate observations from these nonstandard conditional distributions. Due to the complexity of the considered model and data structures, the implementation of the MH algorithm is not straightforward. Some details are also given in Appendix 5.2.

For model comparison, we use $\mathbf{D}_o$ to denote the observed data, which include the directly observable data and the ordered categorical data, and use $\mathbf{D}_m$ to denote the unobservable data, which include latent variables and unobserved data that underlie the ordered categorical data. Moreover, let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}, \boldsymbol{\alpha})$ be the overall unknown parameter vector, where $\boldsymbol{\alpha}$ is a vector that includes all unknown thresholds. Suppose that $\mathbf{D}_o$ has arisen under one of the two competing models M_0 and M_1. For $l = 0, 1$, let $p(\mathbf{D}_o|M_l)$ be the probability density of $\mathbf{D}_o$ under M_l. Recall that the Bayes factor is defined by $B_{10} = p(\mathbf{D}_o|M_1)/p(\mathbf{D}_o|M_0)$. Similarly, a path sampling procedure is presented to compute the Bayes factor for model comparison. Utilizing the idea of data augmentation, $\mathbf{D}_o$ is augmented with $\mathbf{D}_m$ in the computation. Let t be a continuous parameter in [0,1] to link the competing models M_0 and M_1, let $p(\mathbf{D}_m, \mathbf{D}_o|\boldsymbol{\theta}^*, t)$ be the complete-data likelihood, and $U(\boldsymbol{\theta}^*, \mathbf{D}_m, \mathbf{D}_o, t) = d \log p(\mathbf{D}_m, \mathbf{D}_o|\boldsymbol{\theta}^*, t)/dt$. Moreover, let S be the number of fixed grids $\{t_{(s)}\}_{s=1}^S$ between [0, 1] ordered as $0 = t_{(0)} < t_{(1)} < \cdots < t_{(S)} < t_{(S+1)} = 1$. Then, $\log B_{10}$ can be computed as

$$\widehat{\log B_{10}} = \frac{1}{2}\sum_{s=0}^{S}(t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_{(s)}), \tag{5.15}$$

where $\bar{U}_{(s)}$ is the average of the $U(\boldsymbol{\theta}^*, \mathbf{D}_m, \mathbf{D}_o, t)$ on the basis of all simulation draws at $t = t_{(s)}$, that is, $\bar{U}_{(s)} = J^{-1}\sum_{j=1}^J U(\boldsymbol{\theta}^{*(j)}, \mathbf{D}_m^{(j)}, \mathbf{D}_o, t_{(s)})$, in which $\{(\boldsymbol{\theta}^{*(j)}, \mathbf{D}_m^{(j)}),\ j = 1, \cdots, J\}$ are observations simulated from the joint conditional distribution $p(\boldsymbol{\theta}^*, \mathbf{D}_m|\ \mathbf{D}_o, t_{(s)})$.

5.3.4 A Simulation Study

Recall that a dichotomous variable is an ordered categorical variable that is defined by two categories with a threshold of zero. It is usually coded with '0' and '1', and the probability of observing '0' and '1' is decided by an underlying normal distribution with a fixed threshold. Another kind of discrete variable which also has two categories and usually coded with '0' and '1' is the binary variable. Although they have the same coding, binary variables are different from dichotomous variables. Binary variables are unordered, they do not associate with thresholds, and the probabilities of observing '0' and '1' are decided by a binomial distribution rather than the normal distribution. Hence, the methods in analyzing dichotomous variables should not be directly applied to analyze binary variables, otherwise misleading conclusion may be obtained. Given a variable with codings '0' and '1', it is important to decide whether this variable is an ordered dichotomous variable or an unordered binary variable.

Results obtained from a simulation study (see Lee, Song and Cai, 2010) are presented here to illustrate the above idea. In this simulation study, a data set $\mathbf{V} = \{\mathbf{v}_i,\ i = 1, \cdots, n\}$ was generated with $\mathbf{v}_i = (z_{i1}, z_{i2}, z_{i3}, y_{i4}, \cdots, y_{i9})^T$. For $k = 1,\ 2,\ 3,\ z_{ik}$ is dichotomous, while for $k = 4, \cdots, 9,\ y_{ik}$ is binary with a distribution $B(1, p_{ik})$. In the formulation of EFDs, we have $y_{ik} \propto \exp\{y_{ik}\vartheta_{ik} - \log(1 + e^{\vartheta_{ik}})\}$ with $b(\vartheta_{ik}) = \log(1 + e^{\vartheta_{ik}})$ and $\vartheta_{ik} = \log(p_{ik}/(1 - p_{ik}))$. Note that for each of these binomial variables, $\psi_{\epsilon k} = 1.0$ is treated as fixed parameter. Let $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3})^T$ be the latent continuous random vector corresponds to the dichotomous random vector $(z_{i1}, z_{i2}, z_{i3})^T$. We assume that

$$\begin{aligned} y_{ik} &= \mu_k + \mathbf{\Lambda}_k^T \boldsymbol{\omega}_i + \epsilon_i, \quad && k = 1,\ 2,\ 3, \\ \vartheta_{ik} &= \mu_k + \mathbf{\Lambda}_k^T \boldsymbol{\omega}_i, && k = 4, \cdots, 9, \end{aligned}$$

where $\mathbf{\Lambda}_k$ is the kth row of the following loading matrix:

$$\mathbf{\Lambda}^T = \begin{bmatrix} 1 & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{83} & \lambda_{93} \end{bmatrix},$$

where 1's and 0's are treated as fixed for identifying the model. The true values of elements in $\boldsymbol{\mu}$ and λ_{ij} are given by: $\mu_1 = \cdots = \mu_9 = 1.0; \lambda_{21} = \lambda_{31} = 0.7,\ \lambda_{52} = \lambda_{62} = 0.8$, and $\lambda_{83} = \lambda_{93} = 0.7$. The relationships of the latent variables in $\boldsymbol{\omega}_i = (\eta_i, \xi_{i1}, \xi_{i2})^T$ are assessed by the following nonlinear structural equation:

$$\eta_i = bd_i + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1} \xi_{i2} + \delta_i, \tag{5.16}$$

where the fixed covariate d_i is assumed to come from $N[0, 1]$, and the true values for $b, \gamma_1, \gamma_2, \gamma_3$, and ψ_δ are chosen as 0.8, 0.6, 0.6, 0.8, and 0.3, respectively. The true values of ϕ_{11}, ϕ_{21}, and ϕ_{22} in $\mathbf{\Phi}$, the covariance matrix of $(\xi_{i1}, \xi_{i2})^T$, are 1.0, 0.3, and 1.0, respectively. The following prior inputs of the hyperparameter values in the conjugate prior distributions of the parameters are considered: elements in $\mathbf{A}_{0k}$, $\mathbf{\Lambda}_{0k}$, and $\mathbf{\Lambda}_{0\omega k}$ in (5.14) are set equal to the true values; $\mathbf{R}_0^{-1} = 7\mathbf{I}$, $\mathbf{H}_{0k}$, $\mathbf{H}_{0yk}$, and $\mathbf{H}_{0\omega k}$ are taken to be 0.25 times the identity matrices; $\alpha_{0\delta k} = 9$, $\beta_{0\delta k} = 3$, and $\rho_0 = 10$.

The purpose of this simulation study is to address the following questions in practical applications: (i) Does the Bayesian approach produce accurate results for samples with small sizes in analyzing binary or dichotomous data? (ii) What kind of bias will result from incorrectly treating dichotomous variables as binary variables? And vice versa.

The data were created on the basis of the model and the same true values as described above. In fitting the model, three sample sizes $n = 200$, 800, and 2,000 were considered. For each sample size, simulated data were analyzed through the Bayesian approach under the following three cases: (A) correctly treating variables z_{i1}, z_{i2}, and z_{i3} as dichotomous variables, and $y_{i4}, \cdots, y_{i9}$ as binary variables; (B) incorrectly treating all the dichotomous variables as binary variables; and (C) incorrectly treating all the binary variables as dichotomous variables. Results were obtained on the basis of 100 replications.

The simulation studies were conducted using R2WinBUGS (Sturtz, Ligges and Gelman, 2005). The codes of WinBUGS and R2WinBUGS in conducting the current simulation study are given in Appendices 5.3 and 5.4 (see also Lee, Song and Cai, 2010). A few test runs were used to check convergence in order to decide the number of burn-in iterations required for achieving convergence. Results indicated that 15,000 burn-in iterations are sufficient. Hence, the Bayesian results were obtained from 20,000 simulated observations after 15,000 burn-in iterations.

Let $\theta(r)$ be the rth element of $\boldsymbol{\theta}$, $\theta_0(r)$ be the true value of $\theta(r)$, and $\hat{\theta}_j(r)$ be the estimate of $\theta(r)$ at the jth replication. Based on 100 replications, we obtain $\mathrm{M}(\hat{\theta}(r))$ and MC-SD$(\hat{\theta}(r))$, the mean and the standard deviation of the parameter estimates of $\theta(r)$, respectively; and EST-SD$(\hat{\theta}(r))$, the mean of the standard error estimates computed through the square root of $\widehat{\text{Var}}(\boldsymbol{\theta}|\mathbf{V})$. The results are assessed through the following summary statistics:

$$\text{Absolute bias (AB) of } \hat{\theta}(r) = |\text{M}(\hat{\theta}(r)) - \theta_0(r)|,$$

$$\text{Root mean square (RMS) of } \hat{\theta}(r) = \{100^{-1}\sum_{j=1}^{100}[\hat{\theta}_j(r) - \theta_0(r)]^2\}^{1/2},$$

$$\text{SE/SD of } \hat{\theta}(r) = \frac{\text{EST-SD}(\hat{\theta}(r))}{\text{MC-SD}(\hat{\theta}(r))}.$$

Results corresponding to Case (A) and sample sizes of 200, 800, and 2,000 are reported in Table 5.2. We observe that all AB values are quite small. However, for $n = 200$, the RMS values corresponding to parameters λ_{52}, λ_{62}, λ_{83}, λ_{93}, ϕ_{11}, and ϕ_{22} are larger than 0.200. It is expected that the empirical performance would be worse with smaller sample sizes. Results on SE/SD indicate that the standard error estimates overestimate the standard deviation of the estimates, even with $n = 800$.

Table 5.2 here

Based on these simulation results, we have the following conclusions in analyzing nonlinear SEMs with binary and/or dichotomous variables: (i) It requires comparatively larger sample sizes than the analysis with continuous variables. To analyze a model with 23 unknown parameters, a sample size of about 800 is required to produce accurate result. (ii) Except for situations with large sample sizes, the standard error estimates obtained from the square roots of $\widehat{\text{Var}}(\boldsymbol{\theta}|\mathbf{V})$ overestimate the true standard deviations. The commonly used z-score that depends on these standard error estimates should not be used in hypothesis testing.

To study whether binary variables can be treated as ordinal variables, and vice versa, the same data sets were reanalyzed under Cases (B) and (C). Results obtained under

Cases (B), (C), and Type I prior inputs are presented in Tables 5.3 and 5.4. From Table 5.3, the AB and/or RMS values corresponding to $\mu_1, \mu_2, \mu_3, \gamma_1, \gamma_2, \gamma_3, b, \phi_{11}$, and ϕ_{22} are quite large. These parameters are related to the dichotomous variables z_{i1}, z_{i2}, and z_{i3} that are incorrectly modeled as binary data through a logit link. From Table 5.4, we observe that the AB and/or RMS values corresponding to $\mu_4, \cdots, \mu_9, \lambda_{52}, \lambda_{62}, \lambda_{83}, \lambda_{93}$, γ_3, ϕ_{11}, and ϕ_{22} are quite large. These parameters are related to the binary variables $y_{i4}, \cdots, y_{i9}$ that are incorrectly modeled as dichotomous data through a threshold specification. The general empirical performances are not improved with increase of sample size. Again, standard error estimates overestimate the true standard deviations. Clearly, it can be concluded that incorrectly treating binary data as dichotomous one, or vice versa would produce misleading results.

Tables 5.3 and 5.4 here

5.4 SEMs with Missing Data

5.4.1 Introduction

Missing data are very common in substantive research. For example, missing data arise from situations where respondents in a household survey may refuse to report income, or individuals in an opinion survey may refuse to express their attitudes toward some sensitive or embarrassing questions. Clearly, observations with missing entries need to be considered for better statistical inferences. In structural equation modeling, much attention has been devoted to analyze models in the presence of missing data. For example, in analyzing standard SEMs, Arbuckle (1996) proposed a full information ML method which maximizes the casewise likelihood of observed continuous data. Recently, Bayesian methods for analyzing missing data in the context of more complex SEMs have been developed. For instance, Song and Lee (2002), and Lee and Song (2004) developed Bayesian methods for analyzing linear and nonlinear SEMs with mixed continuous and ordered categorical variables, on the basis of ignorable missing data that are missing at random (MAR); while Lee and Tang (2006), and Song and Lee (2007) developed Bayesian methods for analyzing nonlinear SEMs with nonignorable missing data.

The main objective of this section is to introduce the Bayesian approach for analyzing SEMs with ignorable missing data that are missing at random (MAR), and nonignorable missing data that are missing with a nonignorable missing mechanism. According to Little and Rubin (2002), the missing data are regarded as MAR if the probability of missingness depends on the fully observed data but not on the missing data themselves. For nonignorable missing data, the probability of missingness depends not only on the observed data but also on the missing data, according to a nonignorable missing model. In the Bayesian approach, we employ the useful strategy that combines the idea of data augmentation and application of MCMC methods. We will show that Bayesian methods for analyzing SEMs with fully observed data can be extended to handle missing data without much theoretical and practical difficulties. Here, we regard observations with missing entries as partially observed data. We will present a Bayesian framework for analyzing general SEMs with missing data that are MAR, including Bayesian estimation, and model comparison via the Bayes factor. Then, we present Bayesian methods to analyze nonlinear SEMs with missing data that are missing with a nonignorable mechanism. Finally, we demonstrate the use of WinBUGS to obtain the Bayesian solutions.

5.4.2 SEMs with Missing Data that are MAR

Let $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$ be a matrix of random vectors with $\mathbf{v}_i = (\mathbf{x}_i^T, \mathbf{y}_i^T)^T$, in which $\mathbf{x}_i$ and $\mathbf{y}_i$ are vectors of continuous variables whose exact measurements are observable and unobservable, respectively. Let $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ be observable ordered categorical data (or dichotomous data) that correspond to $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$. Suppose $\mathbf{v}_i$ follows a general SEM with a vector of latent variables $\boldsymbol{\omega}_i$. Let $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ contain all the latent variables in the model. To deal with the missing data that are MAR, let $\mathbf{X}_{obs}$ and $\mathbf{X}_{mis}$ be the observed and missing data sets corresponding to the continuous data $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$; $\mathbf{Z}_{obs}$ and $\mathbf{Z}_{mis}$ be the observed and missing data sets corresponding to the ordered categorical data $\mathbf{Z}$; $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{mis}$ be the hypothetical observed and missing data sets of $\mathbf{Y}$ corresponding to $\mathbf{Z}_{obs}$ and $\mathbf{Z}_{mis}$, respectively. Moreover, let $\mathbf{V}_{obs} = \{\mathbf{X}_{obs}, \mathbf{Y}_{obs}\}$ and $\mathbf{V}_{mis} = \{\mathbf{X}_{mis}, \mathbf{Y}_{mis}\}$. The main goal is to develop Bayesian methods for estimating the unknown parameter and threshold vector $\boldsymbol{\theta}^* = (\boldsymbol{\theta}, \boldsymbol{\alpha})$ of the model, and comparing competitive models on the basis of the observed data $\{\mathbf{X}_{obs}, \mathbf{Z}_{obs}\}$.

We first consider the Bayesian estimation by investigating the following posterior distribution of $\boldsymbol{\theta}^*$ with given $\mathbf{X}_{obs}$ and $\mathbf{Z}_{obs}$,

$$p(\boldsymbol{\theta}^* | \mathbf{X}_{obs}, \mathbf{Z}_{obs}) \propto p(\mathbf{X}_{obs}, \mathbf{Z}_{obs} | \boldsymbol{\theta}^*) p(\boldsymbol{\theta}^*),$$

where $p(\mathbf{X}_{obs}, \mathbf{Z}_{obs} | \boldsymbol{\theta}^*)$ is the observed-data likelihood and $p(\boldsymbol{\theta}^*)$ is the prior density of $\boldsymbol{\theta}^*$. As $p(\boldsymbol{\theta}^* | \mathbf{X}_{obs}, \mathbf{Z}_{obs})$ is usually very complicated, we utilize the idea of data augmentation (Tanner and Wong, 1987), and then perform the posterior simulation with the MCMC methods. Naturally, the observed data $\{\mathbf{X}_{obs}, \mathbf{Z}_{obs}\}$ are augmented with the latent and missing quantities $\{\boldsymbol{\Omega}, \mathbf{X}_{mis}, \mathbf{Y}_{mis}, \mathbf{Y}_{obs}\} = \{\boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}\}$. As $\mathbf{Y}_{mis}$ is included, it is not necessary to augment with the corresponding $\mathbf{Z}_{mis}$. A sufficiently large number of random observations will be simulated from the joint posterior distribution $[\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs} | \mathbf{X}_{obs}, \mathbf{Z}_{obs}]$. This task can be completed by a hybrid algorithm that combines the Gibbs sampler (Geman and Geman, 1984) and the MH algorithm (Metropolis *et al.*, 1953; Hastings, 1970) as before. Bayesian estimates of parameters in $\boldsymbol{\theta}^*$ and the standard error estimates can be obtained through the sample of simulated observations, $\{(\boldsymbol{\theta}^{*(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{V}_{mis}^{(j)}, \mathbf{Y}_{obs}^{(j)}),\ j = 1, \cdots, J\}$, that are drawn from the joint posterior distribution.

The conditional distributions $p(\boldsymbol{\theta} | \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs})$, $p(\boldsymbol{\Omega} | \boldsymbol{\theta}^*, \mathbf{V}_{mis}, \mathbf{Y}_{obs}, \mathbf{X}_{obs},$

$\mathbf{Z}_{obs}$), $p(\mathbf{V}_{mis}|\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs})$, and $p(\boldsymbol{\alpha}, \mathbf{Y}_{obs}|\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{X}_{obs}, \mathbf{Z}_{obs})$ are required in the implementation of the Gibbs sampler. With $\mathbf{Y} = (\mathbf{Y}_{mis}, \mathbf{Y}_{obs})$ and $\mathbf{V} = (\mathbf{V}_{mis}, \mathbf{V}_{obs})$ given, the conditional distributions corresponding to $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$ can be derived in exactly the same way as in the situation with fully observed data. Similarly, with $\boldsymbol{\theta}$, $\boldsymbol{\Omega}$, $\mathbf{V}_{mis}$, $\mathbf{X}_{obs}$, and $\mathbf{Z}_{obs}$ given, the conditional distribution corresponding to $(\boldsymbol{\alpha}, \mathbf{Y}_{obs})$ can be similarly derived as before. We only need to derive the conditional distribution corresponding to $\mathbf{V}_{mis}$. Under the usual mild assumption that $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are mutually independent, it follows that:

$$p(\mathbf{V}_{mis}|\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs}) = \prod_{i=1}^{n} p(\mathbf{v}_{i,mis}|\boldsymbol{\theta}^*, \boldsymbol{\omega}_i, \mathbf{y}_{i,obs}, \mathbf{x}_{i,obs}, \mathbf{z}_{i,obs}), \tag{5.17}$$

where $\mathbf{v}_{i,mis} = (\mathbf{x}_{i,mis}, \mathbf{y}_{i,mis})$ is the ith data point in the random sample of size n. The individual $\mathbf{v}_{i,mis}$ can be separately simulated from the conditional distribution in (5.17). For most SEMs, the conditional distribution $p(\mathbf{v}_{i,mis}|\boldsymbol{\theta}^*, \boldsymbol{\omega}_i, \mathbf{y}_{i,obs}, \mathbf{x}_{i,obs}, \mathbf{z}_{i,obs})$ is usually simple. Consequently, the computational burden for sampling $\mathbf{V}_{mis}$ is light.

To address the model comparison problem, we let M_0 and M_1 be two competing models, and consider the computation of the following Bayes factor

$$B_{10} = \frac{p(\mathbf{X}_{obs}, \mathbf{Z}_{obs}|M_1)}{p(\mathbf{X}_{obs}, \mathbf{Z}_{obs}|M_0)}.$$

The $\log B_{10}$ can be similarly computed with path sampling (Gelman and Meng, 1998). Consider the following class of densities defined by a continuous parameter t in $[0, 1]$:

$$p(\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}|\mathbf{X}_{obs}, \mathbf{Z}_{obs}, t) = p(\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs}|t)/z(t).$$

where $z(t) = p(\mathbf{X}_{obs}, \mathbf{Z}_{obs}|t)$. The parameter t connects the competing models M_0 and M_1 such that $z(1) = p(\mathbf{X}_{obs}, \mathbf{Z}_{obs}|t = 1) = p(\mathbf{X}_{obs}, \mathbf{Z}_{obs}|M_1)$ and $z(0) = p(\mathbf{X}_{obs}, \mathbf{Z}_{obs}|t = 0) = p(\mathbf{X}_{obs}, \mathbf{Z}_{obs}|M_0)$, and $B_{10} = z(1)/z(0)$. Based on the reasoning given before, it can be similarly shown that

$$\widehat{\log B}_{10} = \frac{1}{2}\sum_{s=0}^{S}(t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_{(s)}), \tag{5.18}$$

where $0 = t_{(0)} < t_{(1)} < \cdots < t_{(S)} < t_{(S+1)} = 1$, which are fixed grids at $[0, 1]$, and

$$\bar{U}_{(s)} = J^{-1} \sum_{j=1}^{J} U(\boldsymbol{\theta}^{*(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{V}_{mis}^{(j)}, \mathbf{Y}_{obs}^{(j)}, \mathbf{X}_{obs}, \mathbf{Z}_{obs}, t_{(s)}), \tag{5.19}$$

in which $\{(\boldsymbol{\theta}^{*(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{V}_{mis}^{(j)}, \mathbf{Y}_{obs}^{(j)}),\ j = 1, \cdots, J\}$ is a sample of observations simulated from $p(\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}|\mathbf{X}_{obs}, \mathbf{Z}_{obs}, t_{(s)})$, and

$$U(\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs}, t) = \frac{d}{dt} \log p(\boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs}|\boldsymbol{\theta}^*, t),$$

where $p(\boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs}|\boldsymbol{\theta}^*, t)$ is the complete-data likelihood. Because a program for simulating $(\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs})$ has already been developed in the Bayesian estimation, the implementation of the path sampling procedure does not require too much extra programming effort. Basically, the Bayesian methodologies for analyzing SEMs with fully observed data can be generalized to handle missing data that are MAR with only one additional simple component in the Gibbs sampler; see (5.17).

For instance, consider the linear SEM with mixed continuous and ordered categorical variables, see Equations (5.1) and (5.2). Since $\boldsymbol{\Psi}_\epsilon$ is diagonal, $\mathbf{v}_{i,mis}$ is independent of $\mathbf{v}_{i,obs} = (\mathbf{x}_{i,obs}, \mathbf{y}_{i,obs})$. Let p_i be the dimension of $\mathbf{v}_{i,mis}$, it follows from (5.1) that

$$p(\mathbf{V}_{mis}|\mathbf{X}_{obs}, \mathbf{Z}_{obs}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Y}_{obs}) = \prod_{i=1}^{n} p(\mathbf{v}_{i,mis}|\boldsymbol{\theta}, \boldsymbol{\omega}_i), \quad \text{and}$$

$$[\mathbf{v}_{i,mis}|\boldsymbol{\theta}, \boldsymbol{\omega}_i] \stackrel{D}{=} N[\boldsymbol{\mu}_{i,mis} + \boldsymbol{\Lambda}_{i,mis}\boldsymbol{\omega}_i, \boldsymbol{\Psi}_{i,mis}],$$

where $\boldsymbol{\mu}_{i,mis}$ is a $p_i \times 1$ subvector of $\boldsymbol{\mu}$ with elements corresponding to observed components deleted, $\boldsymbol{\Lambda}_{i,mis}$ is a $p_i \times q$ submatrix of $\boldsymbol{\Lambda}$ with rows corresponding to observed components deleted, and $\boldsymbol{\Psi}_{i,mis}$ is a $p_i \times p_i$ submatrix of $\boldsymbol{\Psi}_\epsilon$ with the appropriate rows and columns deleted. Hence, even the form of $\mathbf{V}_{mis}$ is complicated with many distinct missing patterns, its conditional distribution only involves a product of very simple univariate normal distributions.

5.4.3 An Illustrative Example

To illustrate the methodology, a portion of the data set obtained from a study (Morisky *et al.*, 1998) of the effects of establishment policies, knowledge, and attitudes on condom use among Filipino commercial sex workers (CSWs) is analyzed. As commercial sex work promotes the spread of AIDS and other sexually transmitted diseases, promotion of safer sexual practice among CSWs is an important issue. In this example, we assume that there are no 'establishment' effects, so that observations obtained within the establishment are identically and independently distributed. A more subtle two-level SEM will be introduced in Section 6.2 to relax this assumption. The data set was collected from female CSWs in 97 establishments (bars, night clubs, etc) in cities of Philippines. The entire questionnaire consists of 134 items, covering the areas of demographics knowledge, attitudes, beliefs, behaviors, self-efficacy for condom use, and social desirability. In our illustrative example, only six observed variables $(v_1, \cdots, v_6)$ are selected. Variables v_1 and v_2 are related to 'worry about getting AIDS', v_3 and v_4 are related to 'aggressiveness', while v_5 and v_6 are about 'attitude to the risk of getting AIDS'. Variables v_3 and v_4 are continuous, while the others are ordered categorical measured with a 5-point scale. We assume that the missing values are missing at random. After deleting obvious outliers, the data set contains 1080 observations, only 754 of them are fully observed. The missing patterns are presented in Table 5.5. Note that some missing patterns only have a very small number of observations. To unify scales of the continuous variables, the corresponding continuous data are standardized. The sample means and sample standard deviations of the continuous variables are $\{1.58, 1203.74\}$ and $\{1.84, 1096.32\}$, respectively. The cell frequencies of each individual ordered categorical variable range from 21 to 709. See Morisky *et al.* (1998) for other descriptive statistics.

Table 5.5 here

To identify parameters associated with the ordered categorical variables, α_{11}, α_{14}, α_{21}, α_{24}, α_{31}, α_{34}, α_{41}, and α_{44} are fixed at -0.478, 1.034, -1.420, 0.525, -0.868, 0.559, -2.130, and -0.547, respectively. These fixed values are calculated via $\alpha_{kh} = \Phi^{*^{-1}}(f_{kh})$, where $\Phi^*(\cdot)$ is the distribution function of $N[0, 1]$, and f_{kh} are observed cumulative proportions of the categories with $z_k < h$. Based on the meanings of the questions corresponding to the selected observed variables, the data are analyzed through a model with three latent variables η, ξ_1, and ξ_2, together with the measurement equation as specified in (5.1), where $\boldsymbol{\mu} = \mathbf{0}$, and

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

Competing models associated with the following different structural equations but the same measurement equation are considered for illustrative purpose:

$$\begin{aligned} M_1: &\quad \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \delta, \\ M_2: &\quad \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1^2 + \delta, \\ M_3: &\quad \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_4\xi_1\xi_2 + \delta, \\ M_4: &\quad \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_5\xi_2^2 + \delta. \end{aligned}$$

Note that M_1 is nested in M_2, M_3, and M_4, whilst M_2, M_3, and M_4 are nonnested. Estimated logarithm Bayes factors are obtained using the path sampling procedure with $S = 20$ and $J = 1,000$. Assuming that we have no prior information from other sources, we conduct an initial Bayesian estimation based on M_1 with noninformative priors in order to get prior inputs of some hyperparameters. Here, we fixed $\mathbf{H}_{0yk} = \mathbf{I}$ and $\mathbf{H}_{0\omega k} = \mathbf{I}$; for other prior inputs, we take $\rho_0 = 10$, $\mathbf{R}_0^{-1} = 4\mathbf{I}$, $\alpha_{0\epsilon k} = \alpha_{0\delta k} = 8$, $\beta_{0\epsilon k} = \beta_{0\delta k} = 10$; and $\boldsymbol{\Lambda}_{0k}$ and $\boldsymbol{\Lambda}_{0\omega k}$ equal to the estimates of $\boldsymbol{\Lambda}_k$ and $\boldsymbol{\Lambda}_{\omega k}$ in the preliminary analysis.

We are interested in comparing the linear model M_1 with the nonlinear models. It is easy to construct a path to link the competing models. For example, the link model M_t for M_1 and M_2 is $M_t: \ \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + t\gamma_3\xi_1^2 + \delta$. Hence, when $t = 0$, $M_t = M_1$; and when $t = 1$, $M_t = M_2$. We obtain the following results: $\{\widehat{\log B}_{21}, \widehat{\log B}_{31}, \widehat{\log B}_{41}\} = \{2.303, 0.340, 0.780\}$. In comparing M_2 and M_1, $\widehat{\log B}_{21} = 2.303$ clearly recommends the nonlinear model M_2. From $\widehat{\log B}_{31}$ and $\widehat{\log B}_{41}$, we see that the other nonlinear models are not significantly better than M_1. Hence, M_2 is the best model among $M_1, \cdots, M_4$.

To compare M_2 with more complex models, we consider the following models with more complicated structural equations:

$$M_5: \quad \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1^2 + \gamma_4\xi_1\xi_2 + \delta,$$

$$M_6: \quad \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1^2 + \gamma_5\xi_2^2 + \delta.$$

Here, M_2 is nested in M_5 and M_6. The estimated logarithm Bayes factors are $\widehat{\log B}_{52} = 0.406$ and $\widehat{\log B}_{62} = 0.489$. We see that the more complex models are not significantly better than M_2; hence the simpler model M_2 is selected. We compute that the PP p-value (Gelman, Meng and Stern, 1996) corresponding to M_2 is 0.572. This indicates that M_2 fits the data well. For completeness, the estimates of unknown parameters in M_2 are reported in Table 5.6.

Table 5.6 here

Based on the results obtained, a nonlinear SEM has been chosen. Its specification about $\mathbf{\Lambda}$ in the measurement equation suggests that there are three non-overlapping latent factors η, ξ_1, and ξ_2, which can be roughly interpreted as 'worry about AIDS', 'aggressiveness' of CSWs, and 'attitude to the risk of getting AIDS'. These latent factors are related through the following nonlinear structural equation: $\eta = 0.544\xi_1 - 0.033\xi_2 - 0.226\xi_1^2$. Thus, 'aggressiveness' of the CSWs has both linear and quadratic effects on 'worry about

AIDS'. Plotting the quadratic curve of η against ξ_1, we find that the maximum of η is roughly at $\xi_1 = 1.2$, and η decreases as ξ_1 moves away from both directions at 1.2. From the model comparison results, the model with the quadratic term of 'attitude to the risk of getting AIDS' or the corresponding interaction term with 'aggressiveness' is not as good. Thus, these nonlinear relationships are not important, and considering the more complicated model that involve both the interaction and quadratic terms is not necessary.

5.4.4 Nonlinear SEMs with Nonignorable Missing Data

Many missing data in behavioral, medical, social, and psychological research are nonignorable in the sense that the missing data depend on the observed data and the missing data themselves. For example, the side effects of the treatment may make the patients worse and thereby affect patients' participation. Nonignorable missing data, which are more difficult to handle than ignorable missing data, have received a considerable attention in statistics; see Ibrahim, Chen and Lipsitz (2001), among others. In the field of SEM, not much work has been done for analyzing nonignorable missing data. In this subsection, we present a Bayesian approach (see Lee and Tang, 2006) for analyzing a nonlinear SEM with nonignorable missing data. For brevity, we focus on continuous data. However, the Bayesian development can be extended to ordered categorical data, and/or data from exponential family distributions based on the key ideas presented in this section and previous sections, see also Song and Lee (2007). Again, the idea of data augmentation and the MCMC tools will be used to obtain the Bayesian estimates of the unknown parameters and the Bayes factor for model comparison. Although Ibrahim, Chen and Lipsitz (2001) pointed out that the parametric form of the assumed missing mechanism itself is not 'testable' from the data, the Bayes factor provides a useful statistic for comparing different missing data models. Moreover, under the context of a given nonignorable missing data model, the Bayes factor can be used to select a better nonlinear SEM for fitting the data.

For each $p \times 1$ random vector $\mathbf{v}_i = (v_{i1}, \cdots, v_{ip})^T$ in the data set $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$, we define a missing indicator $\mathbf{r}_i = (r_{i1}, \cdots, r_{ip})^T$ such that $r_{ij} = 1$ if v_{ij} is missing, and $r_{ij} = 0$ if v_{ij} is observed. Let $\mathbf{r} = (\mathbf{r}_1, \cdots, \mathbf{r}_n)$; and let $\mathbf{V}_{mis}$ and $\mathbf{V}_{obs}$ be the missing and observed data, respectively. If the distribution of $\mathbf{r}$ is independent of $\mathbf{V}_{mis}$, the missing mechanism is defined to be MAR; otherwise the missing mechanism is nonignorable (Little and Rubin, 1987). For nonignorable missing mechanism, it is necessary to investigate the conditional probability of $\mathbf{r}$ given $\mathbf{V}$, $p(\mathbf{r}|\mathbf{V}, \boldsymbol{\varphi})$, where $\boldsymbol{\varphi}$ is a parameter vector. If this probability does not contain unknown parameters in $\boldsymbol{\varphi}$, the missing mechanism is called ignorable and known, otherwise it is called nonignorable and unknown. An example of a nonignorable and unknown mechanism is related to censored data with a known censoring point. For analyzing missing data with a nonignorable and unknown mechanism, the basic issues are specifying a reasonable model for $\mathbf{r}$ given $\mathbf{V}$, and developing statistical methods for analyzing this model together with the model in relation to $\mathbf{V}$.

Let $\mathbf{v}_i = (\mathbf{v}_{i,obs}^T, \mathbf{v}_{i,mis}^T)^T$, where $\mathbf{v}_{i,obs}$ is a $p_{i1} \times 1$ vector of observed components, $\mathbf{v}_{i,mis}$ is a $p_{i2} \times 1$ vector of missing components, and $p_{i1} + p_{i2} = p$. Here, we assume an arbitrary pattern of missing data in $\mathbf{v}_i$, and thus $\mathbf{v}_i = (\mathbf{v}_{i,obs}^T, \mathbf{v}_{i,mis}^T)^T$ may represent some permutation of the indices of the original $\mathbf{v}_i$. Let $[\mathbf{r}_i|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}]$ be the conditional distribution of $\mathbf{r}_i$ given $\mathbf{v}_i$ and $\boldsymbol{\omega}_i$ with a probability density function $p(\mathbf{r}_i|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})$. Let $\boldsymbol{\theta}$ be the structural parameter vector that contains all unknown distinct parameters in $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\Psi}_\delta$, and $\boldsymbol{\Phi}$. The main interest is on the posterior analyses of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ based on $\mathbf{r}$ and $\mathbf{V}_{obs} = \{\mathbf{v}_{i,obs},\ i = 1, \cdots, n\}$. According to the definition of the model, the joint posterior density of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ based on $\mathbf{V}_{obs}$ and $\mathbf{r}$ is given by:

$$p(\boldsymbol{\theta}, \boldsymbol{\varphi}|\mathbf{V}_{obs}, \mathbf{r}) \propto \left\{ \prod_{i=1}^{n} \int_{\boldsymbol{\omega}_i, \mathbf{v}_{i,mis}} p(\mathbf{v}_i|\boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\mathbf{r}_i|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}) p(\boldsymbol{\omega}_i|\boldsymbol{\theta}) d\boldsymbol{\omega}_i d\mathbf{v}_{i,mis} \right\} p(\boldsymbol{\theta}, \boldsymbol{\varphi}),$$

where $p(\boldsymbol{\theta}, \boldsymbol{\varphi})$ denotes the joint prior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$. In general, the above integral does not have a closed form and its dimension is equal to the sum of the dimensions of $\boldsymbol{\omega}_i$ and $\mathbf{v}_{i,mis}$.

We consider the selection of a model for the nonignorable missing mechanism. Theoretically, any general model can be taken. However, a too complex model will induce difficulty in deriving the corresponding conditional distributions of the missing responses given the observed data, and inefficient sampling from those conditional distributions. Now, since the observations are independent,

$$p(\mathbf{r}|\mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\varphi}) = \prod_{i=1}^{n} p(\mathbf{r}_i|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}),$$

where $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$. As the covariance matrix of the error measurement $\boldsymbol{\epsilon}_i$ is diagonal, it follows that when $\boldsymbol{\omega}_i$ is given, the components of $\mathbf{v}_i$ are independent. Hence, for $j \neq l$, it is reasonable to assume that the conditional distributions of r_{ij} and r_{il} given $\boldsymbol{\omega}_i$ are independent. Under this assumption, we use the following binomial model for the nonignorable missing mechanism (see Ibrahim, Chen and Lipsitz, 2001; Lee and Tang, 2006; Song and Lee, 2007):

$$p(\mathbf{r}|\mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\varphi}) = \prod_{i=1}^{n}\prod_{j=1}^{p} \{\mathrm{pr}(r_{ij} = 1|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\}^{r_{ij}}\{1 - \mathrm{pr}(r_{ij} = 1|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\}^{1-r_{ij}}. \quad (5.20)$$

Ibrahim, Chen and Lipsitz (2001) pointed out that as r_{ij} is binary, one can use a sequence of logistic regressions for modeling $\mathrm{pr}(r_{ij} = 1|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})$ in (5.20). They also pointed out that this model has the potential for reducing the number of parameters in the missing data mechanism, yields correlation structures between the r_{ij}'s, allows more flexibility in specifying the missing data model, and facilitates efficient sampling from the conditional distribution of the missing response given the observed data. Following their suggestion, the following logistic regression model is used:

$$\begin{aligned} &\mathrm{logit}\{\mathrm{pr}(\mathbf{r_{ij}} = \mathbf{1}|\mathbf{v_i}, \boldsymbol{\omega}_\mathbf{i}, \boldsymbol{\varphi})\} \\ &= \varphi_0 + \varphi_1 v_{i1} + \cdots + \varphi_p v_{ip} + \varphi_{p+1}\omega_{i1} + \cdots + \varphi_{p+q}\omega_{iq} = \boldsymbol{\varphi}^T\mathbf{e}_i, \end{aligned} \quad (5.21)$$

where $\mathbf{e}_i = (1, v_{i1}, \cdots, v_{ip}, \omega_{i1}, \cdots, \omega_{iq})^T$, and $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \cdots, \varphi_{p+q})^T$. On the basis of the definition of the measurement equation and the basic concepts of latent variables and their indicators in SEMs, in which the characteristics of $\boldsymbol{\omega}_i$ are revealed by the observed variables in $\mathbf{v}_i$, it may be desirable to adopt the following special case of (5.21) which does not depend on $\boldsymbol{\omega}_i$:

$$\mathrm{logit}\{\mathrm{pr}(\mathbf{r_{ij}} = \mathbf{1}|\mathbf{v_i}, \boldsymbol{\omega}_\mathbf{i}, \boldsymbol{\varphi})\} = \varphi_\mathbf{0} + \varphi_\mathbf{1}\mathbf{v_{i1}} + \cdots + \varphi_\mathbf{p}\mathbf{v_{ip}} \qquad (5.22)$$

However, for generality, the Bayesian approach will be presented on the basis of the more general model (5.21). As (5.22) is equivalent to specifying $\varphi_{p+1} = \cdots = \varphi_{p+q} = 0$ in (5.21), modifications of the general Bayesian development in handling this special case are straightforward. We do not recommend to routinely use (5.21) or (5.22) for modeling the nonignorable missing mechanism in every practical application. Other missing mechanism model may be preferable for situations where one is certain about its specific form for the missing mechanism. However, the model specified in (5.21) or (5.22) is reasonable, and is useful for sensitivity analysis of the estimates with respect to missing data with other different missing mechanisms.

Let $\boldsymbol{\Lambda}_k^T$ and $\boldsymbol{\Lambda}_{\omega k}^T$ be the kth rows of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_\omega$, respectively; $\psi_{\epsilon k}$ and $\psi_{\delta k}$ be the kth diagonal elements of $\boldsymbol{\Psi}_\epsilon$ and $\boldsymbol{\Psi}_\delta$, respectively. Let $\mathbf{V}_{mis} = \{\mathbf{v}_{i,mis},\ i = 1, \cdots, n\}$ be the set of missing values associated with the observed variables. The observed data $\mathbf{V}_{obs}$ and the missing data indicator $\mathbf{r}$ are augmented with the missing quantities $\{\mathbf{V}_{mis}, \boldsymbol{\Omega}\}$ in the posterior analysis. The Gibbs sampler (Geman and Geman, 1984) is used to generate a sequence of random observations from the joint posterior distribution $[\boldsymbol{\Omega}, \mathbf{V}_{mis}, \boldsymbol{\theta}, \boldsymbol{\varphi}|\mathbf{V}_{obs}, \mathbf{r}]$, and the Bayesian estimates are then obtained from the observations of this generated sequence. Specifically, observations $\{\boldsymbol{\Omega}, \mathbf{V}_{mis}, \boldsymbol{\theta}, \boldsymbol{\varphi}\}$ are sampled iteratively from the following conditional distributions: $p(\boldsymbol{\Omega}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}) = p(\boldsymbol{\Omega}|\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r})$, $p(\mathbf{V}_{mis}|\ \mathbf{V}_{obs}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r})$, $p(\boldsymbol{\varphi}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{r}) = p(\boldsymbol{\varphi}|\mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{r})$, and

$p(\boldsymbol{\theta}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \boldsymbol{\Omega}, \boldsymbol{\varphi}, \mathbf{r}) = p(\boldsymbol{\theta}|\mathbf{V}, \boldsymbol{\Omega})$. Note that because of the data augmentation, the last conditional distribution does not depend on $\mathbf{r}$, and can be obtained as before. Details of these conditional distributions are presented in Appendix 5.5.

We again use the Bayes factor for model comparison. Let M_0 and M_1 be two competing models, the Bayes factor is defined as

$$B_{10} = \frac{p(\mathbf{V}_{obs}, \mathbf{r}|M_1)}{p(\mathbf{V}_{obs}, \mathbf{r}|M_0)},$$

where

$$p(\mathbf{V}_{obs}, \mathbf{r}|M_k) = \int p(\mathbf{V}_{obs}, \mathbf{r}|\boldsymbol{\theta}_k, \boldsymbol{\varphi}_k)p(\boldsymbol{\theta}_k, \boldsymbol{\varphi}_k)d\boldsymbol{\theta}_k d\boldsymbol{\varphi}_k, \quad k = 0,\ 1,$$

is the marginal density of M_k with parameter vectors $\boldsymbol{\theta}_k$ and $\boldsymbol{\varphi}_k$, and $p(\boldsymbol{\theta}_k, \boldsymbol{\varphi}_k)$ is the prior density of $\boldsymbol{\theta}_k$ and $\boldsymbol{\varphi}_k$. The logarithm Bayes factor is computed using the path sampling procedure (Gelman and Meng, 1998) as follows. Let $U(\boldsymbol{\theta}, \boldsymbol{\varphi}, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{V}_{obs}, \mathbf{r}, t) = d \log p(\boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{V}_{obs}, \mathbf{r}|\boldsymbol{\theta},\ \boldsymbol{\varphi}, t)/dt$, where $p(\boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{V}_{obs}, \mathbf{r}|\boldsymbol{\theta}, \boldsymbol{\varphi}, t)$ is the complete-data likelihood function defined with a continuous parameter $t \in [0, 1]$. Let $0 = t_{(0)} < t_{(1)} < t_{(2)} < \cdots < t_{(S)} < t_{(S+1)} = 1$ be fixed and ordered grids; $\log B_{10}$ is then estimated by

$$\widehat{\log B}_{10} = \frac{1}{2}\sum_{s=0}^{S}(t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_s),$$

where

$$\bar{U}_{(s)} = J^{-1}\sum_{j=1}^{J} U(\boldsymbol{\theta}^{(j)}, \boldsymbol{\varphi}^{(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{V}_{mis}^{(j)}, \mathbf{V}_{obs}, \mathbf{r}, t_{(s)}),$$

and $\{(\boldsymbol{\theta}^{(j)}, \boldsymbol{\varphi}^{(j)}, \boldsymbol{\Omega}^{(j)}, \mathbf{V}_{mis}^{(j)}),\ j = 1, \cdots, J\}$ are observations that are simulated from $p(\boldsymbol{\theta}, \boldsymbol{\varphi},\ \boldsymbol{\Omega}, \mathbf{V}_{mis}|\mathbf{V}_{obs}, \mathbf{r}, t_{(s)})$.

5.4.5 An Illustrative Real Example

To give an illustration of the Bayesian methodology, 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 analyzed in this example, see also Lee (2007). Here, eight variables in original data set (variables 116, 117, 252, 253, 254, 296, 298, 314, see Appendix 1.1) are taken as observed variables in $\mathbf{v} = (v_1, \cdots, v_8)$. These variables are measured on a 10-point scale, for convenience, they are treated as continuous. We choose the data corresponding to females in Russia, who either answered question 116 or 117, or both. Under this choice, most of the data were obtained from working females. There are 712 random observations in the data set in which there are only 451 (63.34%) fully observed cases. The missing data are rather complicated, with 69 different missing patterns. Considering that the questions are either related to personal attitudes or related to personal morality, the corresponding missing data are treated as nonignorable. To unify the scales roughly, the raw data are standardized using the sample mean and sample standard deviation obtained from the fully observed data.

Based on the meanings of the questions corresponding to the observed variables, we use a nonlinear SEM with the following specifications. For the measurement equation, we consider $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_8)^T$, and the following factor loading matrix with a non-overlapping structure:

$$\boldsymbol{\Lambda}^T = \begin{bmatrix} 1 & \lambda_{21} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda_{42} & \lambda_{52} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{73} & \lambda_{83} \end{bmatrix},$$

which corresponds to latent variables η, ξ_1 and ξ_2. The 1's and 0's in $\boldsymbol{\Lambda}$ are fixed to identify the model and to achieve a clear interpretation of latent variables. The latent variable η can be roughly interpreted as 'job satisfaction', and the latent variables ξ_1 and ξ_2 can be roughly interpreted as 'job attitude' and 'morality (in relation to money)', respectively. We first consider the following model M_1, which involves an encompassing structural equation with all second order terms of ξ_{i1} and ξ_{i2}:

$$M_1: \quad \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i1}\xi_{i2} + \gamma_4\xi_{i1}^2 + \gamma_5\xi_{i2}^2 + \delta_i.$$

The following three models are considered for assessing the missing data in this example:

$$\begin{aligned}
M_a: &\quad \mathrm{logit}\{\mathrm{pr}(r_{ij}=1|\mathbf{v}_i,\boldsymbol{\omega}_i,\boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 v_{i1} + \cdots + \varphi_8 v_{i8}, \\
M_b: &\quad \mathrm{logit}\{\mathrm{pr}(r_{ij}=1|\mathbf{v}_i,\boldsymbol{\omega}_i,\boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 \eta_i + \varphi_2 \xi_{i1} + \varphi_3 \xi_{i2}, \\
M_c: &\quad \text{MAR}.
\end{aligned}$$

Note that M_a involves all the observed variables, while M_b involves all the latent variables.

The logarithm Bayes factors for comparing the above models M_a, M_b, and M_c under M_1 are computed via the path sampling procedure. The prior inputs in the conjugate prior distributions are selected as before via an auxiliary estimation. The number of grids in the path sampling procedure for computing all the logarithm Bayes factors is taken to be 10; and for each $t_{(s)}$, 5,000 simulated observations collected after 5,000 burn-in iterations are used to compute $\bar{U}_{(s)}$. The estimated logarithm Bayes factors are equal to $\widehat{\log B^1_{ab}} = 47.34$, and $\widehat{\log B^1_{ac}} = 43.85$, where the superscript of B indicates M_1, and subscripts of B indicate the competing models for the missing mechanism. Clearly, based on these result, the data give strong evidence to support the missing data model M_a, which is in the form of (5.22), for modeling the nonignorable missing data. In addition to the encompassing M_1, we also consider the following nonlinear SEMs:

$$\begin{aligned}
M_2: &\quad \eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1}^2 + \delta_i, \\
M_3: &\quad \eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1}\xi_{i2} + \delta_i, \\
M_4: &\quad \eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i2}^2 + \delta_i.
\end{aligned}$$

Under each of the above models, we compare the missing mechanism models M_a, M_b, and M_c. The estimated logarithm Bayes factors under these models, together with those under M_1, are reported in Table 5.7, for example, $\widehat{\log B^2_{ab}} = 48.56$ and $\widehat{\log B^3_{ac}} = 44.04$. It is clear from the results in this table that for every M_r, $r = 1, \cdots, 4$, the data strongly support M_a. In each case, Bayesian estimates obtained under M_a are significantly different from those obtained under MAR. To save space, these estimates are not reported. Based on the above comparison results, the model M_a is used for comparing M_1, M_2, M_3 and M_4. In this model comparison, the estimated logarithm Bayes factors are equal to $\widehat{\log B_{12}} = -1.29$, $\widehat{\log B_{32}} = -2.59$, and $\widehat{\log B_{42}} = -1.51$. These results give evidence that the data support M_2.

Table 5.7 here

The Bayesian estimates (EST) and their standard error estimates (SE) of the unknown parameters in the selected model M_2 are presented in the left columns of Table 5.8. We observe that the estimates of the coefficients φ_0, φ_2, φ_3, φ_4, φ_5, φ_6, and φ_8 are significantly different from zero. Hence, the nonignorable missing data model that accounts for the nature of the missing data is necessary. The factor loading estimates indicate strong associations between the latent variables and their indicators. From $\hat{\phi}_{11}$, $\hat{\phi}_{12}$ and $\hat{\phi}_{22}$, the estimate of the correlation between ξ_1 and ξ_2 is 0.163. This estimate indicates that 'job attitude, ξ_1' and 'morality, ξ_2' are weakly correlated. The estimated nonlinear structural equation is equal to

$$\eta = -0.103\xi_1 + 0.072\xi_2 + 0.306\xi_1^2.$$

The interpretation of this equation in relation to the effects of the explanatory latent variables ξ_1 and ξ_2 on the outcome latent variable η is similar to the interpretation of other nonlinear structural equation.

Table 5.8 here

The software WinBUGS (Spiegelhalter *et al.*, 2003) can be used to produce Bayesian solutions for SEMs with missing data that are MAR or missing with a nonignorable mechanism. As nonignorable missing data subsume MAR missing data, we focus on the discussion of nonignorable missing data. In writing the WinBUGS code, in addition to specifying the SEM of interest, we need to specify the nonignorable missing data model and the prior distributions in relation to the parameters in the missing data model, etc.

The data set in this illustrative example has been reanalyzed by using WinBUGS with the same settings, for example, the same model structure and same prior inputs. Results obtained on the basis of M_2 with the missing data model M_a are reported in the right columns of Table 5.8. The DIC value corresponding to this model is 16961.2. We observe that the estimates obtained from WinBUGS are close to the estimates that are presented in the left columns of Table 5.8. However, the numerical standard error (SE) estimates produced by this general software are larger than those produced by our tailor-made program for the specific SEM. The WinBUGS code and the data are given in following websites:

http://www.sta.cuhk.edu.hk/song-lee/book-chapter5(section5.4.5)/WinBUGS-code;
http://www.sta.cuhk.edu.hk/song-lee/book-chapter5(section5.4.5)/WinBUGS-data.
(PLEASE CHANGE TO WEB-SITES HOUSED IN JOHN-WILEY).

Appendix 5.1: Conditional Distributions and Implementation of MH Algorithm Related to SEMs with Continuous and Ordered Categorical Variables

We first consider the conditional distribution in Step (a) of the Gibbs sampler. We note that as the underlying continuous measurements in $\mathbf{Y}$ are given, $\mathbf{Z}$ gives no additional information to this conditional distribution. Moreover, as $\mathbf{v}_i$ are conditionally independent, and $\boldsymbol{\omega}_i$ are also conditionally independent among themselves and independent of $\mathbf{Z}$, we have

$$p(\boldsymbol{\Omega}|\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}) = \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\mathbf{v}_i, \boldsymbol{\theta}).$$

It can be shown that

$$[\boldsymbol{\omega}_i|\mathbf{v}_i, \boldsymbol{\theta}] \stackrel{D}{=} N[\boldsymbol{\Sigma}^* \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1}(\mathbf{v}_i - \boldsymbol{\mu}), \boldsymbol{\Sigma}^*], \tag{5.A1}$$

in which $\boldsymbol{\Sigma}^* = (\boldsymbol{\Sigma}_\omega^{-1} + \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1} \boldsymbol{\Lambda})^{-1}$, where $\boldsymbol{\Pi}_0 = \mathbf{I} - \boldsymbol{\Pi}$, and

$$\boldsymbol{\Sigma}_\omega = \begin{bmatrix} \boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^T + \boldsymbol{\Psi}_\delta)\boldsymbol{\Pi}_0^{-T} & \boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}^T\boldsymbol{\Pi}_0^{-T} & \boldsymbol{\Phi} \end{bmatrix},$$

is the covariance matrix of $\boldsymbol{\omega}_i$. An alternative expression for this conditional distribution can be obtained by the following result, $p(\boldsymbol{\omega}_i|\mathbf{v}_i, \boldsymbol{\theta}) \propto p(\mathbf{v}_i|\boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i|\boldsymbol{\xi}_i, \boldsymbol{\theta}) p(\boldsymbol{\xi}_i|\boldsymbol{\theta})$. Based on the definition of the model and assumptions, $p(\boldsymbol{\omega}_i|\mathbf{v}_i, \boldsymbol{\theta})$ is proportional to

$$\begin{aligned} \exp\Big\{ -\frac{1}{2}\big[&(\mathbf{v}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_\epsilon^{-1} (\mathbf{v}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) \\ &+ (\boldsymbol{\eta}_i - \boldsymbol{\Pi}\boldsymbol{\eta}_i - \boldsymbol{\Gamma}\boldsymbol{\xi}_i)^T \boldsymbol{\Psi}_\delta^{-1} (\boldsymbol{\eta}_i - \boldsymbol{\Pi}\boldsymbol{\eta}_i - \boldsymbol{\Gamma}\boldsymbol{\xi}_i) + \boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i\big]\Big\}. \end{aligned} \tag{5.A2}$$

Based on the practical experience available so far, simulating observations on the basis of (5.A1) or (5.A2) give similar and acceptable results for statistical inference.

To derive the conditional distributions with respect to the structural parameters in Step (b), let $\boldsymbol{\theta}_v$ be the unknown parameters in $\boldsymbol{\mu}, \boldsymbol{\Lambda}$, and $\boldsymbol{\Psi}_\epsilon$ associated with (5.1), and let $\boldsymbol{\theta}_\omega$ be the unknown parameters in $\mathbf{\Lambda}_\omega, \mathbf{\Phi}$, and $\mathbf{\Psi}_\delta$ associated with (5.2). It is natural to take prior distributions such that $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_v)p(\boldsymbol{\theta}_\omega)$.

We first consider the conditional distributions corresponding to $\boldsymbol{\theta}_v$. Similar as before, the following commonly used conjugate type prior distributions are used:

$$\boldsymbol{\mu} \stackrel{D}{=} N[\boldsymbol{\mu}_0, \mathbf{\Sigma}_0], \quad \psi_{\epsilon k}^{-1} \stackrel{D}{=} Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}],$$

$$[\mathbf{\Lambda}_k|\psi_{\epsilon k}] \stackrel{D}{=} N[\mathbf{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0vk}], \quad k = 1, \cdots, p,$$

where $\psi_{\epsilon k}$ is the kth diagonal element of $\mathbf{\Psi}_\epsilon$, $\mathbf{\Lambda}_k^T$ is a $1 \times l_k$ row vector that only contains the unknown parameters in the kth row of $\mathbf{\Lambda}$; $\alpha_{0\epsilon k}$, $\beta_{0\epsilon k}$, $\boldsymbol{\mu}_0$, $\mathbf{\Lambda}_{0k}$, $\mathbf{H}_{0vk}$, and $\mathbf{\Sigma}_0$ are hyperparameters whose values are assumed to be given. For $k \neq h$, it is assumed that $(\psi_{\epsilon k}, \mathbf{\Lambda}_k)$ and $(\psi_{\epsilon h}, \mathbf{\Lambda}_h)$ are independent. To cope with the case with fixed known elements in $\mathbf{\Lambda}$, let $\mathbf{L} = (l_{kj})_{p\times q}$ be the index matrix such that $l_{kj} = 0$ if λ_{kj} is known and $l_{kj} = 1$ if λ_{kj} is unknown, and $l_k = \sum_{j=1}^q l_{kj}$. Let $\mathbf{\Omega}_k$ be a submatrix of $\mathbf{\Omega}$ such that the jth row with $l_{kj} = 0$ deleted, and let $\mathbf{v}_k^* = (v_{1k}^*, \cdots, v_{nk}^*)^T$ with

$$v_{ik}^* = v_{ik} - \mu_k - \sum_{j=1}^q \lambda_{kj}\omega_{ij}(1 - l_{kj}),$$

where v_{ik} is the kth element of $\mathbf{v}_i$, and μ_k is the kth element of $\boldsymbol{\mu}$. Let $\mathbf{\Sigma}_{vk} = (\mathbf{H}_{0vk}^{-1} + \mathbf{\Omega}_k\mathbf{\Omega}_k^T)^{-1}, \boldsymbol{\mu}_{vk} = \mathbf{\Sigma}_{vk}[\mathbf{H}_{0vk}^{-1}\mathbf{\Lambda}_{0k} + \mathbf{\Omega}_k\mathbf{v}_k^*]$, and $\beta_{\epsilon k} = \beta_{0\epsilon k} + 2^{-1}(\mathbf{v}_k^{*T}\mathbf{v}_k^* - \boldsymbol{\mu}_{vk}^T\mathbf{\Sigma}_{vk}^{-1}\boldsymbol{\mu}_{vk} + \mathbf{\Lambda}_{0k}^T\mathbf{H}_{0vk}^{-1}\mathbf{\Lambda}_{0k})$. Then, it can be shown that for $k = 1, \cdots, p$,

$$\begin{aligned}
&[\psi_{\epsilon k}^{-1}|\boldsymbol{\mu}, \mathbf{V}, \mathbf{\Omega}] \stackrel{D}{=} Gamma[n/2 + \alpha_{0\epsilon k}, \beta_{\epsilon k}],\\
&[\mathbf{\Lambda}_k|\psi_{\epsilon k}, \boldsymbol{\mu}, \mathbf{V}, \mathbf{\Omega}] \stackrel{D}{=} N[\boldsymbol{\mu}_{vk}, \psi_{\epsilon k}\mathbf{\Sigma}_{vk}],\\
&[\boldsymbol{\mu}|\mathbf{\Lambda}, \mathbf{\Psi}_\epsilon, \mathbf{V}, \mathbf{\Omega}] \stackrel{D}{=} N[(\mathbf{\Sigma}_0^{-1} + n\mathbf{\Psi}_\epsilon^{-1})^{-1}(n\mathbf{\Psi}_\epsilon^{-1}\widetilde{\mathbf{V}} + \mathbf{\Sigma}_0^{-1}\boldsymbol{\mu}_0),\ (\mathbf{\Sigma}_0^{-1} + n\mathbf{\Psi}_\epsilon^{-1})^{-1}],
\end{aligned} \tag{5.A3}$$

where $\widetilde{\mathbf{V}} = \sum_{i=1}^n (\mathbf{v}_i - \mathbf{\Lambda}\boldsymbol{\omega}_i)/n$.

Now, consider the conditional distribution of $\boldsymbol{\theta}_\omega$. As the parameters in $\boldsymbol{\theta}_\omega$ are only involved in the structural equation, this conditional distribution is proportional to $p(\mathbf{\Omega}|\boldsymbol{\theta}_\omega)$

$p(\boldsymbol{\theta}_\omega)$, which is independent of $\mathbf{V}$ and $\mathbf{Z}$. Let $\boldsymbol{\Omega}_1 = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_n)$ and $\boldsymbol{\Omega}_2 = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n)$. Since the distribution of $\boldsymbol{\xi}_i$ only involves $\boldsymbol{\Phi}$, $p(\boldsymbol{\Omega}_2|\boldsymbol{\theta}_\omega) = p(\boldsymbol{\Omega}_2|\boldsymbol{\Phi})$. Moreover, we take the prior distribution of $\boldsymbol{\Phi}$ such that it is independent of the prior distributions of $\boldsymbol{\Lambda}_\omega$ and $\boldsymbol{\Psi}_\delta$. It follows that $p(\boldsymbol{\Omega}|\boldsymbol{\theta}_\omega)p(\boldsymbol{\theta}_\omega) \propto [p(\boldsymbol{\Omega}_1|\boldsymbol{\Omega}_2, \boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\delta)p(\boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\delta)][p(\boldsymbol{\Omega}_2|\boldsymbol{\Phi})p(\boldsymbol{\Phi})]$. Hence, the marginal conditional densities of $(\boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\delta)$ and $\boldsymbol{\Phi}$ can be treated separately.

Consider a conjugate type prior distribution for $\boldsymbol{\Phi}$ with $\boldsymbol{\Phi}^{-1} \stackrel{D}{=} W_{q_2}[\mathbf{R}_0, \rho_0]$, where ρ_0 and the positive definite matrix $\mathbf{R}_0$ are the given hyperparameters. It can be shown that

$$[\boldsymbol{\Phi}|\boldsymbol{\Omega}_2] \stackrel{D}{=} IW_{q_2}[(\boldsymbol{\Omega}_2\boldsymbol{\Omega}_2^T + \mathbf{R}_0^{-1}), n + \rho_0]. \tag{5.A4}$$

Similar as before, the prior distributions of elements in $(\boldsymbol{\Psi}_\delta, \boldsymbol{\Lambda}_\omega)$ are taken as

$$\psi_{\delta k}^{-1} \stackrel{D}{=} Gamma[\alpha_{0\delta k}, \beta_{0\delta k}], \quad [\boldsymbol{\Lambda}_{\omega k}|\psi_{\delta k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0\omega k}, \psi_{\delta k}\mathbf{H}_{0\omega k}],$$

where $k = 1, \cdots, q_1, \boldsymbol{\Lambda}_{\omega k}^T$ is a $1 \times l_{\omega k}$ row vector that contains the unknown parameters in the kth row of $\boldsymbol{\Lambda}_\omega$; $\alpha_{0\delta k}, \beta_{0\delta k}, \boldsymbol{\Lambda}_{0\omega k}$, and $\mathbf{H}_{0\omega k}$ are given hyperparameters. For $h \neq k$, $(\psi_{\delta k}, \boldsymbol{\Lambda}_{\omega k})$ and $(\psi_{\delta h}, \boldsymbol{\Lambda}_{\omega h})$ are assumed to be independent. Let $\mathbf{L}_\omega = (l_{\omega kj})_{q_1 \times q}$ be the index matrix associated with $\boldsymbol{\Lambda}_\omega$, and $l_{\omega k} = \sum_{j=1}^q l_{\omega kj}$. Let $\boldsymbol{\Omega}_k^*$ be the submatrix of $\boldsymbol{\Omega}$ such that all the jth row corresponding to $l_{\omega kj} = 0$ are deleted; and $\boldsymbol{\Omega}_{\eta k}^* = (\eta_{1k}^*, \cdots, \eta_{nk}^*)^T$ with

$$\eta_{ik}^* = \eta_{ik} - \sum_{j=1}^q \lambda_{\omega kj}\omega_{ij}(1 - l_{\omega kj}),$$

where ω_{ij} is the jth element of $\boldsymbol{\omega}_i$. Then, it can be shown that

$$[\psi_{\delta k}^{-1}|\boldsymbol{\Omega}] \stackrel{D}{=} Gamma[n/2 + \alpha_{0\delta k}, \beta_{\delta k}], \quad [\boldsymbol{\Lambda}_{\omega k}|\boldsymbol{\Omega}, \psi_{\delta k}] \stackrel{D}{=} N[\boldsymbol{\mu}_{\omega k}, \psi_{\delta k}\boldsymbol{\Sigma}_{\omega k}], \tag{5.A5}$$

where $\boldsymbol{\Sigma}_{\omega k} = (\mathbf{H}_{0\omega k}^{-1} + \boldsymbol{\Omega}_k^*\boldsymbol{\Omega}_k^{*T})^{-1}$, $\boldsymbol{\mu}_{\omega k} = \boldsymbol{\Sigma}_{\omega k}[\mathbf{H}_{0\omega k}^{-1}\boldsymbol{\Lambda}_{0\omega k} + \boldsymbol{\Omega}_k^*\boldsymbol{\Omega}_{\eta k}^*]$, and $\beta_{\delta k} = \beta_{0\delta k} + 2^{-1}(\boldsymbol{\Omega}_{\eta k}^{*T}\boldsymbol{\Omega}_{\eta k}^* - \boldsymbol{\mu}_{\omega k}^T\boldsymbol{\Sigma}_{\omega k}^{-1}\boldsymbol{\mu}_{\omega k} + \boldsymbol{\Lambda}_{0\omega k}^T\mathbf{H}_{0\omega k}^{-1}\boldsymbol{\Lambda}_{0\omega k})$.

Finally, we consider the joint conditional distribution of $(\boldsymbol{\alpha}, \mathbf{Y})$ given $\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{X}$, and $\mathbf{Z}$. Suppose that the model in relation to the subvector $\mathbf{y}_i = (y_{i1}, \cdots, y_{is})^T$ of $\mathbf{v}_i$ is given by:

$$\mathbf{y}_i = \boldsymbol{\mu}_y + \boldsymbol{\Lambda}_y\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_{yi},$$

where $\boldsymbol{\mu}_y$ $(s \times 1)$ is a subvector of $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}_y$ $(s \times q)$ is a submatrix of $\boldsymbol{\Lambda}$, $\boldsymbol{\epsilon}_{yi}$ $(s \times 1)$ is a subvector of $\boldsymbol{\epsilon}_i$ with diagonal covariance submatrix $\boldsymbol{\Psi}_y$ of $\boldsymbol{\Psi}_\epsilon$. Let $\mathbf{z}_i = (z_{i1}, \cdots, z_{is})^T$ be the ordered categorical observation corresponding to $\mathbf{y}_i$, $i = 1, \cdots, n$. We use the following non-informative prior distribution for the unknown thresholds in $\boldsymbol{\alpha}_k$:

$$p(\alpha_{k,2}, \cdots, \alpha_{k,b_k-1}) \propto C, \quad \text{for} \quad \alpha_{k,2} < \cdots < \alpha_{k,b_k-1}, \quad k = 1, \cdots, s,$$

where C is a constant. Given $\boldsymbol{\Omega}$ and the fact that the covariance matrix $\boldsymbol{\Psi}_y$ is diagonal, the ordered categorical data $\mathbf{Z}$ and the thresholds corresponding to different rows are also conditionally independent. For $k = 1, \cdots, s$, let $\mathbf{Y}_k^T$ and $\mathbf{Z}_k^T$ be the kth rows of $\mathbf{Y}$ and $\mathbf{Z}$, respectively, it can be shown that

$$p(\boldsymbol{\alpha}_k, \mathbf{Y}_k | \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\Omega}) = p(\boldsymbol{\alpha}_k | \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\Omega}) p(\mathbf{Y}_k | \boldsymbol{\alpha}_k, \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\Omega}), \tag{5.A6}$$

with

$$\begin{aligned} p(\boldsymbol{\alpha}_k | \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\Omega}) \propto \prod_{i=1}^{n} \Bigg[& \Phi^* \left\{ \psi_{yk}^{-1/2} (\alpha_{k,z_{ik}+1} - \mu_{yk} - \boldsymbol{\Lambda}_{yk}^T \boldsymbol{\omega}_i) \right\} \\ & - \Phi^* \left\{ \psi_{yk}^{-1/2} (\alpha_{k,z_{ik}} - \mu_{yk} - \boldsymbol{\Lambda}_{yk}^T \boldsymbol{\omega}_i) \right\} \Bigg], \end{aligned} \tag{5.A7}$$

and $p(\mathbf{Y}_k | \boldsymbol{\alpha}_k, \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\Omega})$ is the product of $p(y_{ik} | \boldsymbol{\alpha}_k, \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\Omega})$, where

$$[y_{ik} | \boldsymbol{\alpha}_k, \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\Omega}] \stackrel{D}{=} N[\mu_{yk} + \boldsymbol{\Lambda}_{yk}^T \boldsymbol{\omega}_i, \psi_{yk}] I_{[\alpha_{k,z_{ik}}, \alpha_{k,z_{ik}+1})}(y_{ik}), \tag{5.A8}$$

in which ψ_{yk} is the kth diagonal element of $\boldsymbol{\Psi}_y$, μ_{yk} is the kth element of $\boldsymbol{\mu}_y$, $\boldsymbol{\Lambda}_{yk}^T$ is the kth row of $\boldsymbol{\Lambda}_y$, $I_A(y)$ is an index function which takes 1 if $y \in A$ and 0 otherwise, and $\Phi^*(\cdot)$ denotes the distribution function of $N[0,1]$. As a result,

$$p(\boldsymbol{\alpha}_k, \mathbf{Y}_k | \mathbf{Z}_k, \boldsymbol{\theta}, \boldsymbol{\Omega}) \propto \prod_{i=1}^{n} \phi \left\{ \psi_{yk}^{-1/2} (y_{ik} - \mu_{yk} - \boldsymbol{\Lambda}_{yk}^T \boldsymbol{\omega}_i) \right\} I_{[\alpha_{k,z_{ik}}, \alpha_{k,z_{ik}+1})}(y_{ik}), \tag{5.A9}$$

where $\phi(\cdot)$ is the standard normal density.

To sample from the conditional distributions (5.A2) and (5.A9), the MH algorithm is implemented as follows.

For $p(\boldsymbol{\omega}_i|\mathbf{v}_i,\boldsymbol{\theta})$, we choose $N[\mathbf{0},\sigma^2\boldsymbol{\Sigma}^*]$ as the proposal distribution, where $\boldsymbol{\Sigma}^{*-1} = \boldsymbol{\Sigma}_\omega^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda}$, with

$$\boldsymbol{\Sigma}_\omega^{-1} = \begin{bmatrix} \boldsymbol{\Pi}_0^T\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Pi}_0 & -\boldsymbol{\Pi}_0^T\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Gamma} \\ -\boldsymbol{\Gamma}^T\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Pi}_0 & \boldsymbol{\Phi}^{-1} + \boldsymbol{\Gamma}^T\boldsymbol{\Psi}_\delta^{-1}\boldsymbol{\Gamma} \end{bmatrix}.$$

Let $p(\cdot|\mathbf{0},\sigma^2,\boldsymbol{\Sigma}^*)$ be the proposal density corresponding to $N[\mathbf{0},\sigma^2\boldsymbol{\Sigma}^*]$, where σ^2 is an appropriate preassigned constant. The MH algorithm is implemented as follows: At the jth MH iteration with a current value $\boldsymbol{\omega}_i^{(j)}$, a new candidate $\boldsymbol{\omega}_i$ is generated from $p(\cdot|\boldsymbol{\omega}_i^{(j)},\sigma^2,\boldsymbol{\Sigma}^*)$, and accepting this new candidate with the probability

$$\min\left\{1, \frac{p(\boldsymbol{\omega}_i|\mathbf{v}_i,\boldsymbol{\theta})}{p(\boldsymbol{\omega}_i^{(j)}|\mathbf{v}_i,\boldsymbol{\theta})}\right\},$$

where $p(\boldsymbol{\omega}_i|\mathbf{v}_i,\boldsymbol{\theta})$ is given by (5.A2).

For $p(\boldsymbol{\alpha}_k,\mathbf{Y}_k|\mathbf{Z}_k,\boldsymbol{\theta},\boldsymbol{\Omega})$, we use the equality (5.A6) from Cowles (1996) to construct a joint proposal density for $\boldsymbol{\alpha}_k$, and $\mathbf{Y}_k$ in the MH algorithm for generating observations from it. At the jth MH iteration, we generate a vector of thresholds $(\alpha_{k,2},\cdots,\alpha_{k,b_k-1})$ from the following univariate truncated normal distribution:

$$\alpha_{k,z} \overset{D}{=} N[\alpha_{k,z}^{(j)},\sigma_{\alpha_k}^2]I_{(\alpha_{k,z-1},\alpha_{k,z+1}^{(j)}]}(\alpha_{k,z}) \quad \text{for} \quad z=2,\cdots,b_k-1,$$

where $\alpha_{k,z}^{(j)}$ is the current value of $\alpha_{k,z}$ at the jth iteration of the Gibbs sampler, and $\sigma_{\alpha_k}^2$ is an appropriate preassigned constant. Random observations from the above univariate truncated normal are simulated via the algorithm of Roberts (1995). Then, the acceptance probability for $(\boldsymbol{\alpha}_k,\mathbf{Y}_k)$ as a new observation is $\min\{1,R_k\}$, where

$$R_k = \prod_{z=2}^{b_k-1} \frac{\Phi^*\{(\alpha_{k,z+1}^{(j)} - \alpha_{k,z}^{(j)})/\sigma_{\alpha_k}\} - \Phi^*\{(\alpha_{k,z-1} - \alpha_{k,z}^{(j)})/\sigma_{\alpha_k}\}}{\Phi^*\{(\alpha_{k,z+1} - \alpha_{k,z})/\sigma_{\alpha_k}\} - \Phi^*\{(\alpha_{k,z-1}^{(j)} - \alpha_{k,z})/\sigma_{\alpha_k}\}} \times$$
$$\prod_{i=1}^{n} \frac{\Phi^*\left\{\psi_{yk}^{-1/2}(\alpha_{k,z_{ik}+1} - \mu_{yk} - \boldsymbol{\Lambda}_{yk}^T\boldsymbol{\omega}_i)\right\} - \Phi^*\left\{\psi_{yk}^{-1/2}(\alpha_{k,z_{ik}} - \mu_{yk} - \boldsymbol{\Lambda}_{yk}^T\boldsymbol{\omega}_i)\right\}}{\Phi^*\left\{\psi_{yk}^{-1/2}(\alpha_{k,z_{ik}+1}^{(j)} - \mu_{yk} - \boldsymbol{\Lambda}_{yk}^T\boldsymbol{\omega}_i)\right\} - \Phi^*\left\{\psi_{yk}^{-1/2}(\alpha_{k,z_{ik}}^{(j)} - \mu_{yk} - \boldsymbol{\Lambda}_{yk}^T\boldsymbol{\omega}_i)\right\}}.$$

As R_k only depends on the old and new values of $\boldsymbol{\alpha}_k$ and not on $\mathbf{Y}_k$, it does not require to generate a new $\mathbf{Y}_k$ in any iteration in which the new value of $\boldsymbol{\alpha}_k$ is not accepted (see Cowles, 1996). For an accepted $\boldsymbol{\alpha}_k$, a new $\mathbf{Y}_k$ is simulated from (5.A8).

Appendix 5.2: Conditional Distributions and Implementation of MH Algorithm Related to SEMs with EFDs

It can be shown that the full conditional distribution of $\mathbf{\Omega}$ is given by

$$p(\mathbf{\Omega}|\mathbf{Y},\boldsymbol{\theta}) = \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\mathbf{y}_i,\boldsymbol{\theta}) \propto \prod_{i=1}^{n} p(\mathbf{y}_i|\boldsymbol{\omega}_i,\boldsymbol{\theta})p(\boldsymbol{\eta}_i|\boldsymbol{\xi}_i,\boldsymbol{\theta})p(\boldsymbol{\xi}_i|\boldsymbol{\theta}),$$

where $p(\boldsymbol{\omega}_i|\mathbf{y}_i,\boldsymbol{\theta})$ is proportional to

$$\exp\Bigg\{\sum_{k=1}^{p}\Big[y_{ik}\vartheta_{ik} - b(\vartheta_{ik})\Big]/\psi_{\epsilon k} \tag{5.A10}$$
$$-\frac{1}{2}\Big[(\boldsymbol{\eta}_i - \mathbf{B}\mathbf{d}_i - \mathbf{\Pi}\boldsymbol{\eta}_i - \mathbf{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i))^T\mathbf{\Psi}_\delta^{-1}(\boldsymbol{\eta}_i - \mathbf{B}\mathbf{d}_i - \mathbf{\Pi}\boldsymbol{\eta}_i - \mathbf{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i)) + \boldsymbol{\xi}_i^T\mathbf{\Phi}^{-1}\boldsymbol{\xi}_i\Big]\Bigg\}.$$

Under the conjugate prior distributions given in (5.14), it can be shown that the full conditional distributions of the components of $\boldsymbol{\theta}$ are given by

$$p(\mathbf{A}_k|\mathbf{Y},\mathbf{\Omega},\mathbf{\Lambda}_k,\psi_{\epsilon k}) \propto \exp\Bigg\{\sum_{i=1}^{n}\frac{y_{ik}\vartheta_{ik} - b(\vartheta_{ik})}{\psi_{\epsilon k}} - \frac{1}{2}(\mathbf{A}_k - \mathbf{A}_{0k})^T\mathbf{H}_{0k}^{-1}(\mathbf{A}_k - \mathbf{A}_{0k})\Bigg\},$$

$$p(\psi_{\epsilon k}|\mathbf{Y},\mathbf{\Omega},\mathbf{A}_k,\mathbf{\Lambda}_k) \propto \psi_{\epsilon k}^{-(\frac{n}{2}+\alpha_{0\epsilon k}-1)}\exp\Bigg\{\sum_{i=1}^{n}\Big[\frac{y_{ik}\vartheta_{ik} - b(\vartheta_{ik})}{\psi_{\epsilon k}} + c_k(y_{ik},\psi_{\epsilon k})\Big] - \frac{\beta_{0k}}{\psi_{\epsilon k}}\Bigg\},$$

$$p(\mathbf{\Lambda}_k|\mathbf{Y},\mathbf{\Omega},\mathbf{A}_k,\psi_{\epsilon k}) \propto \exp\Bigg\{\sum_{i=1}^{n}\frac{y_{ik}\vartheta_{ik} - b(\vartheta_{ik})}{\psi_{\epsilon k}} - \frac{1}{2}\psi_{\epsilon k}^{-1}(\mathbf{\Lambda}_k - \mathbf{\Lambda}_{0k})^T\mathbf{H}_{0yk}^{-1}(\mathbf{\Lambda}_k - \mathbf{\Lambda}_{0k})\Bigg\},$$

$$[\psi_{\delta k}^{-1}|\mathbf{\Omega},\mathbf{\Lambda}_{\omega k}] \stackrel{D}{=} Gamma[n/2 + \alpha_{0\delta k},\beta_{\delta k}],$$

$$[\mathbf{\Lambda}_{\omega k}|\mathbf{\Omega},\psi_{\delta k}] \stackrel{D}{=} N[\boldsymbol{\mu}_{\omega k},\psi_{\delta k}\mathbf{\Sigma}_{\omega k}],$$

$$[\mathbf{\Phi}|\mathbf{\Omega}] \stackrel{D}{=} IW_{q_2}[(\mathbf{\Omega}_2\mathbf{\Omega}_2^T + \mathbf{R}_0^{-1}), n+\rho_0], \tag{5.A11}$$

where $\mathbf{\Sigma}_{\omega k} = (\mathbf{H}_{0\omega k}^{-1} + \mathbf{G}\mathbf{G}^T)^{-1}$, $\boldsymbol{\mu}_{\omega k} = \mathbf{\Sigma}_{\omega k}(\mathbf{H}_{0\omega k}^{-1}\mathbf{\Lambda}_{0\omega k} + \mathbf{G}\mathbf{\Omega}_{1k})$, and $\beta_{\delta k} = \beta_{0\delta k} + (\mathbf{\Omega}_{1k}^T\mathbf{\Omega}_{1k} - \boldsymbol{\mu}_{\omega k}^T\mathbf{\Sigma}_{\omega k}^{-1}\boldsymbol{\mu}_{\omega k} + \mathbf{\Lambda}_{0\omega k}^T\mathbf{H}_{0\omega k}^{-1}\mathbf{\Lambda}_{0\omega k})/2$, in which $\mathbf{G} = (\mathbf{G}(\boldsymbol{\omega}_1),\cdots,\mathbf{G}(\boldsymbol{\omega}_n))$, $\mathbf{\Omega}_1 = (\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_n)$, $\mathbf{\Omega}_2 = (\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_n)$, and $\mathbf{\Omega}_{1k}^T$ is the kth row of $\mathbf{\Omega}_1$.

In simulating observations from $p(\boldsymbol{\omega}_i|\mathbf{y}_i,\boldsymbol{\theta})$ in (5.A10), we choose $N[\cdot,\sigma_\omega^2\mathbf{\Omega}_\omega]$ as the proposal distribution in the MH algorithm, where $\mathbf{\Omega}_\omega^{-1} = \mathbf{\Sigma}_\omega^* + \mathbf{\Lambda}^T\mathbf{\Psi}_\omega\mathbf{\Lambda}$, in which

$$\mathbf{\Sigma}_\omega^* = \begin{bmatrix} \mathbf{\Pi}_0^T\mathbf{\Psi}_\delta^{-1}\mathbf{\Pi}_0 & -\mathbf{\Pi}_0^T\mathbf{\Psi}_\delta^{-1}\mathbf{\Gamma}\mathbf{\Delta} \\ -\mathbf{\Delta}^T\mathbf{\Gamma}^T\mathbf{\Psi}_\delta^{-1}\mathbf{\Pi}_0 & \mathbf{\Phi}^{-1} + \mathbf{\Delta}^T\mathbf{\Gamma}^T\mathbf{\Psi}_\delta^{-1}\mathbf{\Gamma}\mathbf{\Delta} \end{bmatrix},$$

with $\mathbf{\Pi}_0 = \mathbf{I}_{q_1} - \mathbf{\Pi}$, $\mathbf{\Delta} = (\partial \mathbf{F}(\boldsymbol{\xi}_i)/\partial \boldsymbol{\xi}_i)^T|_{\boldsymbol{\xi}_i=\mathbf{0}}$, and $\mathbf{\Psi}_\omega = \mathrm{diag}(\ddot{b}(\vartheta_{i1})/\psi_{\epsilon 1}, \cdots, \ddot{b}(\vartheta_{ip})/\psi_{\epsilon p})|_{\boldsymbol{\omega}_i=\mathbf{0}}$.

In simulating observations from the conditional distributions $p(\mathbf{A}_k|\mathbf{Y}, \mathbf{\Omega}, \mathbf{\Lambda}_k, \psi_{\epsilon k})$, $p(\psi_{\epsilon k}\, |\mathbf{Y}, \mathbf{\Omega}, \mathbf{A}_k, \mathbf{\Lambda}_k)$, and $p(\mathbf{\Lambda}_k|\mathbf{Y}, \mathbf{\Omega}, \mathbf{A}_k, \psi_{\epsilon k})$, the proposal distributions are $N[\cdot, \sigma_a^2 \mathbf{\Omega}_{ak}]$, $N[\cdot,\, \sigma_\psi^2 \Omega_{\psi k}]$, and $N[\cdot, \sigma_\lambda^2 \mathbf{\Omega}_{\lambda k}]$, respectively, where

$$\mathbf{\Omega}_{ak}^{-1} = \sum_{i=1}^n \ddot{b}(\vartheta_{ik}) \mathbf{c}_{ik} \mathbf{c}_{ik}^T/\psi_{\epsilon k} \Big|_{\mathbf{A}_k=\mathbf{0}} + \mathbf{H}_{0k}^{-1},$$

$$\Omega_{\psi k}^{-1} = 1 - n/2 - \alpha_{0\epsilon k} - 2\sum_{i=1}^n [y_{ik}\vartheta_{ik} - b(\vartheta_{ik})] - \ddot{c}_k(y_{ik}, \psi_{\epsilon k})\Big|_{\psi_{\epsilon k}=1} + 2\beta_{0\epsilon k},$$

$$\mathbf{\Omega}_{\lambda k}^{-1} = \sum_{i=1}^n \ddot{b}(\vartheta_{ik}) \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T \Big|_{\mathbf{\Lambda}_k=\mathbf{0}} + \psi_{\epsilon k}^{-1} \mathbf{H}_{0yk}^{-1}.$$

For improving efficiency, we respectively use $N[\boldsymbol{\mu}_{ak}, \mathbf{\Omega}_{ak}]$, $N[\mu_{\psi k}, \mathbf{\Omega}_{\psi k}]$, and $N[\boldsymbol{\mu}_{\lambda k}, \mathbf{\Omega}_{\lambda k}]$ as initial proposal distributions in the first few iterations, where

$$\boldsymbol{\mu}_{ak} = \sum_{i=1}^n \left[y_{ik} - \dot{b}(\vartheta_{ik})|_{\mathbf{A}_k=\mathbf{0}} \right] \frac{\mathbf{c}_{ik}}{\psi_{\epsilon k}} + \mathbf{H}_{0k}^{-1} \mathbf{A}_{0k},$$

$$\mu_{\psi k} = 1 - n/2 - \alpha_{0\epsilon k} - \sum_{i=1}^n \left[y_{ik}\vartheta_{ik} - b(\vartheta_{ik}) \right] + \dot{c}_k(y_{ik}, \psi_{\epsilon k})\Big|_{\psi_{\epsilon k}=1} + \beta_{0\epsilon k},$$

$$\boldsymbol{\mu}_{\lambda k} = \sum_{i=1}^n \left[y_{ik} - \dot{b}(\vartheta_{ik})|_{\Lambda_k=\mathbf{0}} \right] \frac{\boldsymbol{\omega}_i}{\psi_{\epsilon k}} + \mathbf{H}_{0yk}^{-1} \mathbf{\Lambda}_{0k}.$$

Let $\mathbf{y}_k^{*T}$ be the kth row of $\mathbf{Y}$ that is not directly observable, $\mathbf{z}_k$ be the corresponding ordered categorical vector, and $\boldsymbol{\alpha}_k = (\alpha_{k,2}, \cdots, \alpha_{k,b_k-1})$. It can be shown by similar derivation as in Appendix 5.1 that

$$\begin{aligned} &p(\boldsymbol{\alpha}_k, \mathbf{y}_k^* | \mathbf{z}_k, \mathbf{\Omega}, \boldsymbol{\theta}) = p(\boldsymbol{\alpha}_k | \mathbf{z}_k, \mathbf{\Omega}, \boldsymbol{\theta}) p(\mathbf{y}_k^* | \boldsymbol{\alpha}_k, \mathbf{z}_k, \mathbf{\Omega}, \boldsymbol{\theta}) \propto \\ &\prod_{i=1}^n \exp\Big\{ [y_{ik}^* \vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\epsilon k} + c_k(y_{ik}^*, \psi_{\epsilon k}) \Big\} I_{[\alpha_{k,z_{ik}}, \alpha_{k,z_{ik}+1})}(y_{ik}^*), \end{aligned} \tag{5.A12}$$

where $I_A(y)$ is an indicator function which takes 1 if $y \in A$, and 0 otherwise. The treatment of dichotomous variables is similar.

A multivariate version of the MH algorithm is used to simulate observations from $p(\boldsymbol{\alpha}_k, \mathbf{y}_k^* |\, \mathbf{z}_k, \mathbf{\Omega}, \boldsymbol{\theta})$ in (5.A12). Following Cowles (1996), for the joint proposal distribution of $\boldsymbol{\alpha}_k$ and $\mathbf{y}_k^*$ given $\mathbf{z}_k$, $\boldsymbol{\Omega}$, and $\boldsymbol{\theta}$ can be constructed according to the factorization $p(\boldsymbol{\alpha}_k, \mathbf{y}_k^*|\mathbf{z}_k, \boldsymbol{\Omega}, \boldsymbol{\theta}) = p(\boldsymbol{\alpha}_k|\mathbf{z}_k, \boldsymbol{\Omega}, \boldsymbol{\theta})p(\mathbf{y}_k^*|\boldsymbol{\alpha}_k, \mathbf{z}_k, \boldsymbol{\Omega}, \boldsymbol{\theta})$. At the jth iteration, we generate a candidate vector of thresholds $(\alpha_{k,2}, \cdots, \alpha_{k,b_k-1})$ from the following univariate truncated normal distribution

$$\alpha_{k,m} \sim N[\alpha_{k,m}^{(j)}, \sigma_{\alpha_k}^2]I_{(\alpha_{k,m-1}, \alpha_{k,m+1}^{(j)}]}(\alpha_{k,m}), \quad \text{for} \quad m = 2, \cdots, b_k - 1,$$

where $\alpha_{k,m}^{(j)}$ is the current value of $\alpha_{k,m}$, and $\sigma_{\alpha_k}^2$ is chosen to obtain an average acceptance rate of approximately 0.25 or greater. The acceptance probability for a candidate vector $(\boldsymbol{\alpha}_k, \mathbf{y}_k^*)$ as a new observation $(\boldsymbol{\alpha}_k^{(j+1)}, \mathbf{y}_k^{*(j+1)})$ is $\min\{1, R_k\}$, where

$$R_k = \frac{p(\boldsymbol{\alpha}_k, \mathbf{y}_k^*|\mathbf{z}_k, \boldsymbol{\Omega}, \boldsymbol{\theta})p(\boldsymbol{\alpha}_k^{(j)}, \mathbf{y}_k^{*(j)}|\boldsymbol{\alpha}_k, \mathbf{y}_k^*, \mathbf{z}_k, \boldsymbol{\Omega}, \boldsymbol{\theta})}{p(\boldsymbol{\alpha}_k^{(j)}, \mathbf{y}_k^{*(j)}|\mathbf{z}_k, \boldsymbol{\Omega}, \boldsymbol{\theta})p(\boldsymbol{\alpha}_k, \mathbf{y}_k^*|\boldsymbol{\alpha}_k^{(j)}, \mathbf{y}_k^{*(j)}, \mathbf{z}_k, \boldsymbol{\Omega}, \boldsymbol{\theta})}.$$

For an accepted $\boldsymbol{\alpha}_k$, a new $\mathbf{y}_k^*$ is simulated from the following univariate truncated distribution:

$$[y_{ik}^*|\boldsymbol{\alpha}_k, z_{ik}, \boldsymbol{\omega}_i, \boldsymbol{\theta}] \stackrel{D}{=} \exp\{[y_{ik}^*\vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\epsilon k} + c_k(y_{ik}^*, \psi_{\epsilon k})\}I_{[\alpha_{k,z_{ik}}, \alpha_{k,z_{ik}+1})}(y_{ik}^*),$$

where y_{ik}^* and z_{ik} are the ith components of $\mathbf{y}_k^*$ and $\mathbf{z}_k$, respectively, and $I_A(y)$ is an indicator function which takes 1 if y in A and zero otherwise.

Appendix 5.3: WinBUGS Code Related to Section 5.3.4

```
model {
    for(i in 1:N){
        #Measurement equation model
        for(j in 1:3){
            y[i,j]~dnorm(mu[i,j],1)I(low[z[i,j]+1],high[z[i,j]+1])
        }
        for(j in 4:P){
            z[i,j]~dbin(pb[i,j],1)
            pb[i,j]<-exp(mu[i,j])/(1+exp(mu[i,j]))
        }
        mu[i,1]<-uby[1]+eta[i]
        mu[i,2]<-uby[2]+lam[1]*eta[i]
```

```
        mu[i,3]<-uby[3]+lam[2]*eta[i]
        mu[i,4]<-uby[4]+xi[i,1]
        mu[i,5]<-uby[5]+lam[3]*xi[i,1]
        mu[i,6]<-uby[6]+lam[4]*xi[i,1]
        mu[i,7]<-uby[7]+xi[i,2]
        mu[i,8]<-uby[8]+lam[5]*xi[i,2]
        mu[i,9]<-uby[9]+lam[6]*xi[i,2]

        #Structural equation model
        xi[i,1:2]~dmnorm(zero2[1:2],phi[1:2,1:2])
        eta[i]~dnorm(etamu[i],psd)
        etamu[i]<-ubeta*c[i]+gam[1]*xi[i,1]+gam[2]*xi[i,2]
                  +gam[3]*xi[i,1]*xi[i,2]
    } #End of i

    for(i in 1:2){ zero2[i]<-0 }

    #Priors inputs for loadings and coefficients
    for (i in 1:P){ uby[i]~dnorm(1.0,4.0) }
    lam[1]~dnorm(0.7,4.0);   lam[2]~dnorm(0.7,4.0)
    lam[3]~dnorm(0.8,4.0);   lam[4]~dnorm(0.8,4.0)
    lam[5]~dnorm(0.7,4.0);   lam[6]~dnorm(0.7,4.0)
    ubeta~dnorm(0.8,4.0)

    var.gam<-4.0*psd
    gam[1]~dnorm(0.6,var.gam);   gam[2]~dnorm(0.6,var.gam)
    gam[3]~dnorm(0.8,var.gam)

    #Priors inputs for precisions
    psd~dgamma(9,3);    sgd<-1/psd
    phi[1:2,1:2]~dwish(R[1:2,1:2], 10)
    phx[1:2,1:2]<-inverse(phi[1:2,1:2])
} #end
```

Appendix 5.4: R2WinBUGS Code Related to Section 5.3.4

```
library(MASS)      #Load the MASS package
library(R2WinBUGS) #Load the R2WinBUGS package
library(boa)       #Load the boa package
```

```
N<-2000; P<-9

phi<-matrix(data=c(1.0,0.3,0.3,1.0),ncol=2) #The covariance matrix of xi
Ro<-matrix(data=c(7.0,2.1,2.1,7.0), ncol=2)
yo<-matrix(data=NA,nrow=N,ncol=P); p<-numeric(P); v<-numeric(P)

#Matrices save the Bayesian Estimates and Standard Errors
Eu<-matrix(data=NA,nrow=100,ncol=9)
SEu<-matrix(data=NA,nrow=100,ncol=9)
Elam<-matrix(data=NA,nrow=100,ncol=6)
SElam<-matrix(data=NA,nrow=100,ncol=6)
Egam<-matrix(data=NA,nrow=100,ncol=3)
SEgam<-matrix(data=NA,nrow=100,ncol=3)
Ephx<-matrix(data=NA,nrow=100,ncol=3)
SEphx<-matrix(data=NA,nrow=100,ncol=3)
Eb<-numeric(100); SEb<-numeric(100)
Esgd<-numeric(100); SEsgd<-numeric(100)

#Arrays save the HPD intervals
uby=array(NA, c(100,9,2))
ubeta=array(NA, c(100,2))
lam=array(NA, c(100,6,2))
gam=array(NA, c(100,3,2))
sgd=array(NA, c(100,2))
phx=array(NA, c(100,3,2))

DIC=numeric(100)    #DIC values

#Parameters to be estimated
parameters<-c("uby","ubeta","lam","gam","sgd","phx")

#Initial values for the MCMC in WinBUGS
init1<-list(uby=rep(1.0,9),ubeta=0.8,lam=c(0.7,0.7,0.8,0.8,0.7,0.7),
gam=c(0.6,0.6,0.8),psd=3.33,phi=matrix(data=c(1.0989,-0.3267,-0.3267,
1.0989),ncol=2,byrow=TRUE), xi=matrix(data=rep(0.0,4000),ncol=2))

init2<-list(uby=rep(1.0,9),ubeta=1.0,lam=rep(1.0,6),gam=(1.0,3),
```

```
psd=3.0,phi=matrix(data=c(1.0,0.0,0.0,1.0),ncol=2,byrow=TRUE),
xi=matrix(data=rep(0.0,4000),ncol=2))

inits<-list(init1, init2)

#Do simulation for 100 replications
for (t in 1:100) {
    #Generate the data for the simulation study
    for (i in 1:N) {
        #Generate xi
        xi<-mvrnorm(1,mu=c(0,0),phi)
        #Generate the fixed covariates
        co<-rnorm(1,0,1)
        #Generate error term in structural equation
        delta<-rnorm(1,0,sqrt(0.3))
        #Generate eta according to the structural equation
        eta<-0.8*co[i]+0.6*xi[1]+0.6*xi[2]+0.8*xi[1]*xi[2]+delta
        #Generate error terms in measurement equations
        eps<-rnorm(3,0,1)

        #Generate theta according to measurement equations
        v[1]<-1.0+eta+eps[1]; v[2]<-1.0+0.7*eta+eps[2]
        v[3]<-1.0+0.7*eta+eps[3]
        v[4]<-1.0+xi[1]; v[5]<-1.0+0.8*xi[1]; v[6]<-1.0+0.8*xi[1]
        v[7]<-1.0+xi[2]; v[8]<-1.0+0.7*xi[2]; v[9]<-1.0+0.7*xi[2]

        #transform theta to orinal variables
        for (j in 1:3) { if (v[j]>0) yo[i,j]<-1 else yo[i,j]<-0 }

        #transform theta to binary variables
        for (j in 4:9) {
            p[j]<-exp(v[j])/(1+exp(v[j]))
            yo[i,j]<-rbinom(1,1,p[j])
        }
    }

    #Input data set for WinBUGS
    data<-list(N=2000,P=9,R=Ro,z=yo,c=co,low=c(-2000,0),high=c(0,2000))
```

```
#Call WinBUGS
model<-bugs(data,inits,parameters,model.file="D:/Run/model.txt",
n.chains=2,n.iter=35000,n.burnin=15000,n.thin=1,DIC=True,
bugs.directory="C:/Program Files/WinBUGS14/",
working.directory="D:/Run/")

#Save Bayesian Estimates
Eu[t,]<-model$mean$uby; Elam[t,]<-model$mean$lam;
Egam[t,]<-model$mean$gam
Ephx[t,1]<-model$mean$phx[1,1]; Ephx[t,2]<-model$mean$phx[1,2]
Ephx[t,3]<-model$mean$phx[2,2]; Eb[t]<-model$mean$ubeta
Esgd[t]<-model$mean$sgd

#Save Standard Errors
SEu[t,]<-model$sd$uby; SElam[t,]<-model$sd$lam;
SEgam[t,]<-model$sd$gam
SEphx[t,1]<-model$sd$phx[1,1]; SEphx[t,2]<-model$sd$phx[1,2]
SEphx[t,3]<-model$sd$phx[2,2]; SEb[t]<-model$sd$ubeta
SEsgd[t]<-model$sd$sgd

#Save HPD intervals
for (i in 1:9) {
    temp=model$sims.array[,1,i];
    uby[t,i,]=boa.hpd(temp,0.05)
}
temp=model$sims.array[,1,10]; ubeta[t,]=boa.hpd(temp,0.05)
for (i in 1:6) {
    temp=model$sims.array[,1,10+i];
    lam[t,i,]=boa.hpd(temp,0.05)
}
for (i in 1:3) {
    temp=model$sims.array[,1,16+i];
    gam[t,i,]=boa.hpd(temp,0.05)
}
temp=model$sims.array[,1,20]; sgd[t,]=boa.hpd(temp,0.05)
temp=model$sims.array[,1,21]; phx[t,1,]=boa.hpd(temp,0.05)
temp=model$sims.array[,1,22]; phx[t,2,]=boa.hpd(temp,0.05)
```

```
    temp=model$sims.array[,1,24]; phx[t,3,]=boa.hpd(temp,0.05)

    #Save DIC value
    DIC[t]=model#DIC
}   #end
```

Appendix 5.5: Conditional Distributions Related to SEMs with Nonignorable Missing Data

Consider the conditional distribution $p(\boldsymbol{\Omega}|\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r})$. Note that when $\mathbf{V}_{mis}$ is given, the underlying model with missing data reduces to a nonlinear SEM with fully observed data. Thus, it follows that

$$p(\boldsymbol{\Omega}|\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}) = \prod_{i=1}^{n} p(\boldsymbol{\omega}_i|\mathbf{v}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}_i) \propto \prod_{i=1}^{n} p(\mathbf{v}_i|\boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i|\boldsymbol{\xi}_i, \boldsymbol{\theta}) p(\boldsymbol{\xi}_i|\boldsymbol{\theta}) p(\mathbf{r}_i|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}),$$

where $p(\boldsymbol{\omega}_i|\mathbf{v}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}_i)$ is proportional to

$$\exp\Big\{ -\frac{1}{2}\Big[(\mathbf{v}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_\epsilon^{-1} (\mathbf{v}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\omega}_i) + (\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \mathbf{G}(\boldsymbol{\omega}_i))^T \boldsymbol{\Psi}_\delta^{-1} (\boldsymbol{\eta}_i - \boldsymbol{\Lambda}_\omega \mathbf{G}(\boldsymbol{\omega}_i)) + \boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i\Big] + (\sum_{j=1}^{p} r_{ij}) \boldsymbol{\varphi}^T \mathbf{e}_i - p \log(1 + \exp(\boldsymbol{\varphi}^T \mathbf{e}_i))\Big\}. \tag{5.A13}$$

To derive $p(\mathbf{V}_{mis}|\mathbf{V}_{obs}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r})$, we note that $\mathbf{r}_i$ is independent of $\boldsymbol{\theta}$. Moreover, as $\boldsymbol{\Psi}_\epsilon$ is diagonal, $\mathbf{v}_{i,mis}$ is independent of $\mathbf{v}_{i,obs}$. It can be shown that:

$$p(\mathbf{V}_{mis}|\mathbf{V}_{obs}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}) = \prod_{i=1}^{n} p(\mathbf{v}_{i,mis}|\mathbf{v}_{i,obs}, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}_i) \propto \prod_{i=1}^{n} p(\mathbf{v}_{i,mis}|\boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\mathbf{r}_i|\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}),$$

and $p(\mathbf{v}_{i,mis}|\mathbf{v}_{i,obs}, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}_i)$ is proportional to

$$\frac{\exp\{-\frac{1}{2}(\mathbf{v}_{i,mis} - \boldsymbol{\mu}_{i,mis} - \boldsymbol{\Lambda}_{i,mis}\boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_{i,mis}^{-1} (\mathbf{v}_{i,mis} - \boldsymbol{\mu}_{i,mis} - \boldsymbol{\Lambda}_{i,mis}\boldsymbol{\omega}_i) + (\sum_{j=1}^{p} r_{ij}) \boldsymbol{\varphi}^T \mathbf{e}_i\}}{(1 + \exp(\boldsymbol{\varphi}^T \mathbf{e}_i))^p}, \tag{5.A14}$$

where $\boldsymbol{\mu}_{i,mis}$ is a $p_{i2} \times 1$ subvector of $\boldsymbol{\mu}$ with its elements corresponding to missing components of $\mathbf{v}_i$, $\boldsymbol{\Lambda}_{i,mis}$ is a $p_{i2} \times q$ submatrix of $\boldsymbol{\Lambda}$ with its rows corresponding to missing components of $\mathbf{v}_i$, and $\boldsymbol{\Psi}_{i,mis}$ is a $p_{i2} \times p_{i2}$ submatrix of $\boldsymbol{\Psi}_\epsilon$ with the rows and columns corresponding to missing components of $\mathbf{v}_i$.

Finally, we consider the conditional distribution of $\boldsymbol{\varphi}$ given $\mathbf{V}$, $\boldsymbol{\Omega}$, $\boldsymbol{\theta}$, and $\mathbf{r}$. Let $p(\boldsymbol{\varphi})$ be the prior density of $\boldsymbol{\varphi}$, such that $\boldsymbol{\varphi} \stackrel{D}{=} N[\boldsymbol{\varphi}^0, \boldsymbol{\Sigma}_\varphi]$, where $\boldsymbol{\varphi}^0$ and $\boldsymbol{\Sigma}_\varphi$ are the hyperparameters whose values are assumed to be given by the prior information. Since the distribution of $\mathbf{r}$ only involves $\mathbf{V}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\varphi}$, and it is assumed that the prior distribution of $\boldsymbol{\varphi}$ is independent with the prior distribution of $\boldsymbol{\theta}$, we have

$$p(\boldsymbol{\varphi}|\mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{r}) \propto p(\mathbf{r}|\mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\varphi})p(\boldsymbol{\varphi}).$$

Thus, it follows from (5.20) and (5.21) that $p(\boldsymbol{\varphi}|\mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{r})$ is proportional to

$$\frac{\exp\{\sum_{i=1}^{n}(\sum_{j=1}^{p} r_{ij})\boldsymbol{\varphi}^T\mathbf{e}_i - \frac{1}{2}(\boldsymbol{\varphi} - \boldsymbol{\varphi}^0)^T\boldsymbol{\Sigma}_\varphi^{-1}(\boldsymbol{\varphi} - \boldsymbol{\varphi}^0)\}}{\prod_{i=1}^{n}(1 + \exp(\boldsymbol{\varphi}^T\mathbf{e}_i))^p}. \tag{5.A15}$$

This completes the derivation of the full conditional distributions that are required in the implementation of the Gibbs sampler. The conditional distributions $p(\boldsymbol{\omega}_i|\mathbf{v}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}_i)$, $p(\mathbf{v}_{i,mis}|\mathbf{v}_{i,obs}, \boldsymbol{\omega}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}_i)$ and $p(\boldsymbol{\varphi}|\mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{r})$ are non-standard. Some details of the MH algorithm for simulating observations from these conditional distributions are presented in Appendix 12.1 in Lee (2007).

References

Arbuckle, J. L. (1996) Full information estimation in the presence of incomplete data. In G. A. Marcoulides and R. E. Schumacker (eds.), *Advanced Structural Equation Modeling*, Mahwah, NJ: Lawrence Erlbaum Publishers.

Bock, R. D. and Aitkin, M. (1981) Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, **46**, 443-459.

Booth, J. G. and Hobert, J. P. (1999) Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B*, **61**, 265-285.

Chib, S. and Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, **85**, 347-361.

Cowles, M. K. (1996) Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, **6**, 101-111.

Fayer, P. M. and Hand, D. J. (1997) Factor analysis, causal indicators and quality of life. *Quality of Life Research*, **6**, 139-150.

Fayer, P. M. and Machin, D. (1998) Factor analysis. In M. J. Staquet, R. D. Hayes and P. M. Fayer (eds), *Quality of Life Assessment in Clinical Trials*. New York: Oxford University Press.

Gelman, A. (1996) Inference and monitoring convergence. In W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, pp. 131-144. London: Chapman and Hall.

Gelman, A. and Meng, X. L. (1998) Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, **13**, 163-185.

Gelman, A., Meng, X. L. and Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**, 733-759.

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.

Geyer, C. J. (1992) Practical Markov chain Monte Carlo. *Statistical Science*, **7**, 473-511.

Gibbons, R. D. and Wilcox-Gök, V. (1998) Health service utilization and insurance coverage: a multivariate probit model. *Journal of the American Statistical Association*, **93**, 63-72.

Glonek, G. F. V. and McCullagh, P. (1995) Multivariate logistic models. *Journal of the Royal Statistical Society, Series B*, **57**, 533-546.

Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97-109.

Ibrahim, J. G., Chen, M. H. and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, **88**, 551-564.

Lee, S. Y. (2007) *Structural Equation Modeling: A Bayesian Approach.* UK: John Wiley & Sons, Ltd.

Lee, S. Y., Poon, W. Y. and Bentler, P. M. (1990) Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters*, **9**, 91-97.

Lee, S. Y. and Song, X. Y. (2003) Bayesian analysis of structural equation models with dichotomous variables. *Statistics in Medicine*, **22**, 3073-3088.

Lee, S. Y. and Song, X. Y. (2004) Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, **57**, 131-150.

Lee, S. Y. and Tang, N. S. (2006) Bayesian analysis of nonlinear structural equation models with nonignorable missing data. *Psychometrika*, **71**, 541-564.

Lee, S. Y., Song, X. Y. and Cai, J. H. (2010) A Bayesian approach for nonlinear structural equation models with dichotomous variables using logit and probit links. *Structural Equation Modeling - A Multidisciplinary Journal*, **17**, 280-302.

Lee, S. Y., Song, X. Y., Skevington, S. and Hao, Y. T. (2005) Application of structural equation models to quality of life. *Structural Equation Modeling - A Multidisciplinary Journal.* **12**, 435-453.

Little, R. J. A. and Rubin, D. B. (1987) *Statistical Analysis with Missing Data.* New York: Wiley.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models.* New York: Chapman and Hall.

Meng, X. L. and Schilling, S. (1996) Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*, **91**, 1254-1267.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087-1091.

Morisky, D. E., Tiglao, T. V., Sneed, C. D., Tempongko, S. B., Baltazar, J. C., Detels, R. and Stein, J.A. (1998) The effects of establishment policies, knowledge and attitudes on condom use among Filipina sex workers. *AIDS Care*, **10**, 213-220.

Olschewski, M. and Schumacher, M. (1990) Statistical analysis of quality of life data in cancer clinical trials. *Statistics in Medicine*, **9**, 749-763.

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

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

Roberts, C. P. (1995) Simulation of truncated normal variables. *Statistics and Computing*, **5**, 121-125.

Sammel, M. D., Ryan, L. M. and Legler, J. M. (1997) Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B*, **59**, 667-678.

Shi, J. Q. and Lee, S. Y. (2000) Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society, Series B*, **62**, 77-87.

Song, X. Y. and Lee, S. Y. (2002) Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika*, **67**, 261-288.

Song, X. Y. and Lee, S. Y. (2007) Bayesian analysis of latent variable models with non-ignorable missing outcomes from exponential family. *Statistics in Medicine*, **26**, 681-693.

Spiegelhalter, D. J., Thomas, A., Best, N. G. and Lunn, D. (2003) *WinBUGS User Manual. Version 1.4*. Cambridge, England: MRC Biostatistics Unit.

Staquet, M. J., Hayes, R. D. and Fayes, P. M. (1998) *Quality of Life Assessment in Clinical Trials*. New York: Oxford University Press.

Sturtz, S., Ligges, U. and Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, **12** 1-16.

Tanner, M. A. and Wong, W. H. (1987) The calculation of posterior distributions by data augmentation(with discussion). *Journal of the American statistical Association*, **82**, 528-550.

The WHOQOL Group. (1998) Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*, **28**, 551-558.

World Values Study Group (1994) World Values Survey, 1981-1984 AND 1990-1993. ICPSR version. Ann Arbor, MI: Institute for Social Research (producer). Ann Arbon, MI: Inter-university Consortium for Political and Social Research (distributor).

Table 5.1: Frequencies of the questions in the WHOQOL data set.

	1	2	3	4	5
Q1 Overall QOL	2	34	75	160	67
Q2 Overall health	25	89	71	117	36
Q3 Pain and discomfort	16	49	78	111	84
Q4 Medical treatment dependence	17	48	65	83	125
Q5 Energy and fatigue	16	53	107	86	76
Q6 Mobility	13	33	62	95	135
Q7 Sleep and rest	23	62	73	116	64
Q8 Daily activities	9	55	63	158	53
Q9 Work capacity	19	71	79	116	53
Q10 Positive feeling	8	22	93	165	50
Q11 Spirituality/personal beliefs	8	29	99	137	65
Q12 Memory and concentration	4	22	148	133	31
Q13 Bodily image and appearance	3	30	106	112	87
Q14 Self-esteem	7	38	104	148	41
Q15 Negative feeling	4	35	89	171	39
Q16 Personal relationship	5	16	59	165	93
Q17 Sexual activity	25	48	112	100	53
Q18 Social support	7	6	73	164	88
Q19 Physical safety and security	4	20	147	129	38
Q20 Physical environment	7	20	142	126	43
Q21 Financial resources	15	34	140	87	62
Q22 Daily life information	4	22	102	154	56
Q23 Participation in leisure activity	15	76	102	108	37
Q24 Living condition	4	12	35	173	114
Q25 Health accessibility and quality	4	20	59	205	50
Q26 Transportation	5	16	43	188	86
Total	269	960	2326	3507	1726

Table 5.2: Simulation results corresponding to Case (A).

		$n = 200$			$n = 800$			$n = 2,000$		
Par	True	AB	SE/SD	RMS	AB	SE/SD	RMS	AB	SE/SD	RMS
μ_1	1.0	0.017	1.36	0.137	0.006	1.10	0.100	0.001	1.11	0.069
μ_2	1.0	0.056	1.25	0.150	0.007	0.98	0.096	0.011	1.02	0.061
μ_3	1.0	0.055	1.15	0.163	0.009	1.04	0.092	0.004	1.01	0.059
μ_4	1.0	0.031	1.06	0.163	0.011	1.09	0.087	0.020	1.10	0.060
μ_5	1.0	0.013	1.09	0.164	0.019	1.13	0.090	0.006	0.95	0.067
μ_6	1.0	0.010	1.14	0.155	0.005	1.18	0.083	0.008	0.99	0.065
μ_7	1.0	0.029	1.32	0.133	0.012	1.04	0.092	0.017	0.96	0.068
μ_8	1.0	0.021	1.15	0.152	0.005	1.12	0.085	0.012	0.97	0.064
μ_9	1.0	0.046	1.10	0.166	0.029	1.07	0.093	0.019	1.06	0.061
λ_{21}	0.7	0.105	1.37	0.165	0.041	1.12	0.109	0.019	1.02	0.076
λ_{31}	0.7	0.125	1.34	0.178	0.054	1.04	0.118	0.007	0.98	0.076
λ_{52}	0.8	0.084	1.50	0.234	0.170	1.29	0.226	0.092	1.06	0.166
λ_{62}	0.8	0.083	1.64	0.218	0.092	1.33	0.183	0.099	1.16	0.158
λ_{83}	0.7	0.109	1.58	0.223	0.165	1.20	0.215	0.128	1.04	0.170
λ_{93}	0.7	0.114	1.57	0.224	0.117	1.31	0.183	0.121	1.20	0.151
γ_1	0.6	0.050	1.73	0.128	0.029	1.43	0.106	0.071	1.21	0.101
γ_2	0.6	0.041	1.83	0.120	0.070	1.30	0.127	0.057	1.11	0.106
γ_3	0.8	0.004	2.22	0.112	0.084	1.81	0.128	0.115	1.31	0.149
b	0.8	0.035	1.13	0.137	0.003	1.06	0.083	0.012	1.06	0.059
ϕ_{11}	1.0	0.141	1.76	0.222	0.127	1.48	0.202	0.109	1.21	0.184
ϕ_{12}	0.3	0.007	1.30	0.135	0.055	0.93	0.113	0.085	1.04	0.074
ϕ_{22}	1.0	0.129	1.89	0.208	0.124	1.38	0.216	0.091	1.13	0.191
ψ_δ	0.3	0.180	3.79	0.062	0.129	2.30	0.058	0.088	1.64	0.057

Note: This table and Tables 5.3 and 5.4 are extracted from Lee, Song and Cai (2010).

Table 5.3: Simulation results corresponding to Case (B).

		$n = 200$			$n = 800$			$n = 2,000$		
Par	True	AB	SE/SD	RMS	AB	SE/SD	RMS	AB	SE/SD	RMS
μ_1	1.0	0.287	1.36	0.334	0.429	1.12	0.451	0.572	1.15	0.581
μ_2	1.0	0.492	1.45	0.520	0.613	1.00	0.632	0.689	1.08	0.696
μ_3	1.0	0.490	1.30	0.525	0.616	1.08	0.632	0.663	1.10	0.670
μ_4	1.0	0.033	1.07	0.167	0.001	1.12	0.086	0.011	1.05	0.061
μ_5	1.0	0.004	1.07	0.165	0.008	1.12	0.089	0.001	0.94	0.067
μ_6	1.0	0.026	1.12	0.159	0.006	1.17	0.082	0.000	1.00	0.062
μ_7	1.0	0.029	1.31	0.137	0.003	1.06	0.092	0.007	0.98	0.066
μ_8	1.0	0.009	1.15	0.152	0.006	1.10	0.086	0.006	0.96	0.063
μ_9	1.0	0.032	1.08	0.165	0.018	1.07	0.089	0.013	1.05	0.059
λ_{21}	0.7	0.307	1.62	0.257	0.232	1.21	0.197	0.123	1.07	0.118
λ_{31}	0.7	0.334	1.55	0.278	0.250	1.11	0.213	0.094	1.03	0.104
λ_{52}	0.8	0.002	1.38	0.233	0.074	1.13	0.200	0.027	1.01	0.149
λ_{62}	0.8	0.010	1.58	0.206	0.001	1.19	0.172	0.028	1.10	0.136
λ_{83}	0.7	0.037	1.50	0.212	0.048	1.11	0.184	0.057	0.96	0.153
λ_{93}	0.7	0.037	1.44	0.218	0.003	1.27	0.154	0.051	1.11	0.131
γ_1	0.6	0.136	1.74	0.162	0.315	1.35	0.235	0.565	1.18	0.364
γ_2	0.6	0.124	1.77	0.156	0.352	1.29	0.257	0.529	1.17	0.346
γ_3	0.8	0.144	1.84	0.184	0.358	1.63	0.319	0.588	1.45	0.493
b	0.8	0.390	1.20	0.353	0.467	1.08	0.392	0.605	1.07	0.492
ϕ_{11}	1.0	0.025	1.49	0.245	0.032	1.29	0.213	0.003	1.11	0.177
ϕ_{12}	0.3	0.504	1.19	0.237	0.263	0.89	0.158	0.085	0.98	0.083
ϕ_{22}	1.0	0.052	1.52	0.257	0.068	1.22	0.244	0.022	0.99	0.211
ψ_δ	0.3	0.300	4.01	0.097	0.511	2.26	0.169	1.257	1.46	0.401

Table 5.4: Simulation results corresponding to Case (C).

		$n = 200$			$n = 800$			$n = 2,000$		
Par	True	AB	SE/SD	RMS	AB	SE/SD	RMS	AB	SE/SD	RMS
μ_1	1.0	0.023	1.26	0.133	0.010	1.15	0.085	0.005	1.09	0.063
μ_2	1.0	0.084	1.23	0.163	0.036	0.99	0.097	0.031	1.03	0.064
μ_3	1.0	0.086	1.14	0.175	0.039	1.06	0.094	0.015	1.06	0.055
μ_4	1.0	0.333	1.05	0.352	0.351	1.10	0.356	0.374	1.08	0.376
μ_5	1.0	0.355	0.97	0.374	0.380	1.08	0.384	0.394	0.94	0.396
μ_6	1.0	0.371	1.09	0.386	0.390	1.15	0.393	0.393	0.97	0.394
μ_7	1.0	0.327	1.28	0.340	0.350	1.04	0.355	0.369	0.95	0.372
μ_8	1.0	0.351	1.10	0.366	0.389	1.08	0.392	0.388	0.95	0.390
μ_9	1.0	0.334	1.05	0.351	0.373	1.02	0.377	0.384	1.05	0.386
λ_{21}	0.7	0.158	1.41	0.185	0.115	1.10	0.137	0.073	0.99	0.096
λ_{31}	0.7	0.186	1.33	0.206	0.130	1.06	0.148	0.044	1.02	0.083
λ_{52}	0.8	0.167	1.26	0.261	0.111	1.07	0.194	0.134	1.03	0.162
λ_{62}	0.8	0.167	1.35	0.247	0.181	1.19	0.204	0.128	1.08	0.156
λ_{83}	0.7	0.157	1.38	0.223	0.136	1.14	0.177	0.121	0.97	0.149
λ_{93}	0.7	0.144	1.30	0.230	0.167	1.15	0.185	0.126	1.17	0.136
γ_1	0.6	0.060	1.51	0.147	0.092	1.26	0.135	0.225	1.15	0.175
γ_2	0.6	0.036	1.58	0.138	0.114	1.25	0.145	0.177	1.22	0.148
γ_3	0.8	0.052	2.14	0.130	0.243	1.78	0.234	0.481	1.60	0.405
b	0.8	0.015	1.11	0.130	0.053	1.03	0.091	0.027	1.07	0.060
ϕ_{11}	1.0	0.370	1.73	0.390	0.458	1.37	0.468	0.517	1.25	0.522
ϕ_{12}	0.3	0.428	1.36	0.151	0.523	1.04	0.166	0.561	1.05	0.172
ϕ_{22}	1.0	0.346	2.11	0.362	0.447	1.52	0.457	0.492	1.08	0.501
ψ_δ	0.3	0.246	2.88	0.086	0.304	1.80	0.110	0.392	1.35	0.138

Table 5.5: Missing patterns and their sample sizes: AIDS data set, '$\times$' and 'o' indicate missing and observed entries, respectively.

	Sample	Observed Variables							Sample	Observed Variables					
Pattern	size	1	2	3	4	5	6	Pattern	size	1	2	3	4	5	6
1	784	o	o	o	o	o	o	11	7	×	o	o	o	×	o
2	100	×	o	o	o	o	o	12	7	×	o	o	o	o	×
3	57	o	×	o	o	o	o	13	9	o	×	o	o	×	o
4	6	o	o	×	o	o	o	14	3	o	×	o	o	o	×
5	4	o	o	o	×	o	o	15	1	o	×	o	×	o	o
6	25	o	o	o	o	×	o	16	1	o	×	×	o	o	o
7	26	×	o	o	o	o	×	17	4	o	×	o	o	×	×
8	17	×	×	o	o	o	o	18	2	×	o	o	o	×	×
9	23	o	o	o	o	×	×	19	1	o	o	×	o	×	×
10	2	×	o	×	o	o	o	20	1	×	×	o	o	×	×

Note: This table and the next three tables are extracted from Lee (2007).

Table 5.6: Bayesian Estimates of parameters in M_2.

Par	EST	Par	EST
λ_{21}	0.228	$\psi_{\epsilon 1}$	0.593
λ_{42}	0.353	$\psi_{\epsilon 2}$	0.972
λ_{63}	0.358	$\psi_{\epsilon 3}$	0.519
		$\psi_{\epsilon 4}$	0.943
γ_1	0.544	$\psi_{\epsilon 5}$	0.616
γ_2	-0.033	$\psi_{\epsilon 6}$	1.056
γ_3	-0.226		
		α_{12}	-0.030
ϕ_{11}	0.508	α_{13}	0.340
ϕ_{12}	-0.029	α_{22}	-0.961
ϕ_{22}	0.394	α_{23}	-0.620
		α_{32}	-0.394
ψ_δ	0.663	α_{33}	0.257
		α_{42}	-1.604
		α_{43}	-0.734

Table 5.7: The estimated log Bayes factors: $\log B^r_{ab}$ and $\log B^r_{ac}$, $r = 1, 2, 3, 4$.

SEM	$\log B^r_{ab}$	$\log B^r_{ac}$
M_1	47.34	43.85
M_2	48.56	46.66
M_3	50.40	44.04
M_4	50.38	44.69

Table 5.8: Bayesian estimates and their standard error estimates of M_2 with M_a.

	Our Method		WinBUGS			Our Method		WinBUGS	
Par	EST	SE	EST	SE	Par	EST	SE	EST	SE
φ_0	-2.791	0.043	-2.794	0.076	μ_1	-0.135	0.038	-0.139	0.065
φ_1	0.038	0.033	0.040	0.059	μ_2	-0.136	0.032	-0.129	0.058
φ_2	-0.280	0.037	-0.280	0.068	μ_3	0.018	0.023	0.015	0.039
φ_3	0.370	0.036	0.365	0.073	μ_4	0.004	0.023	0.005	0.041
φ_4	-0.265	0.041	-0.262	0.083	μ_5	-0.129	0.026	-0.139	0.045
φ_5	-0.455	0.070	-0.502	0.126	μ_6	-0.046	0.023	-0.040	0.041
φ_6	-0.405	0.073	-0.341	0.154	μ_7	0.053	0.026	0.045	0.046
φ_7	0.059	0.056	0.013	0.134	μ_8	0.144	0.026	0.141	0.045
φ_8	0.332	0.038	0.323	0.061	λ_{21}	0.917	0.129	0.830	0.168
					λ_{42}	0.307	0.060	0.317	0.123
					λ_{52}	0.328	0.068	0.320	0.119
					λ_{73}	1.244	0.122	0.955	0.203
					λ_{83}	0.455	0.071	0.388	0.114
					$\psi_{\epsilon 1}$	0.544	0.067	0.508	0.096
					$\psi_{\epsilon 2}$	0.637	0.060	0.673	0.080
					$\psi_{\epsilon 3}$	0.493	0.058	0.492	0.111
					$\psi_{\epsilon 4}$	0.935	0.033	0.932	0.059
					$\psi_{\epsilon 5}$	0.907	0.039	0.922	0.068
					$\psi_{\epsilon 6}$	0.640	0.042	0.548	0.095
					$\psi_{\epsilon 7}$	0.612	0.051	0.714	0.086
					$\psi_{\epsilon 8}$	1.065	0.040	1.065	0.069
					γ_1	-0.103	0.047	-0.103	0.085
					γ_2	0.072	0.052	0.044	0.081
					γ_3	0.306	0.083	0.317	0.139
					ϕ_{11}	0.459	0.057	0.459	0.113
					ϕ_{12}	0.062	0.016	0.071	0.033
					ϕ_{22}	0.316	0.041	0.405	0.096
					ψ_δ	0.413	0.056	0.463	0.105

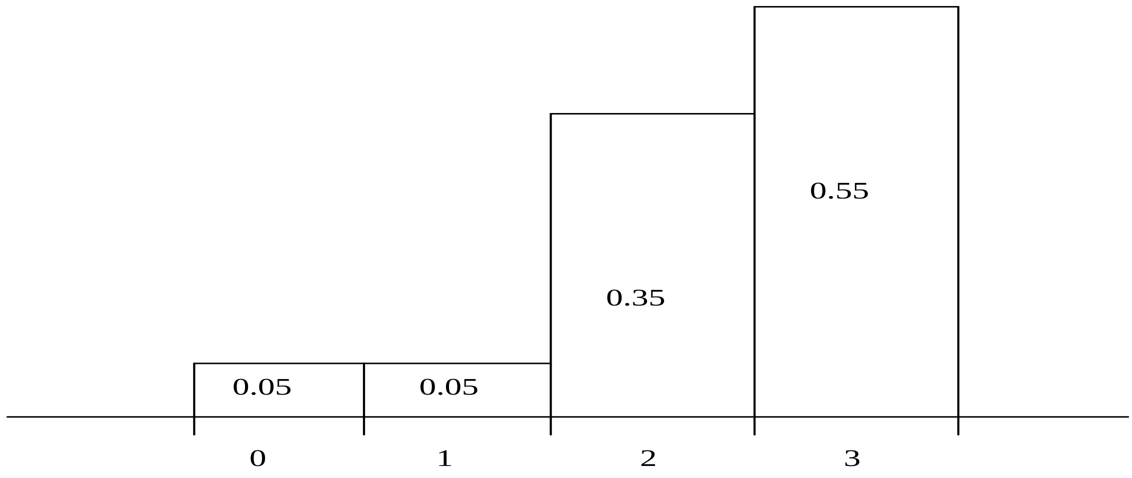

Figure 5.1: Histogram of a hypothetical ordered categorical data set.

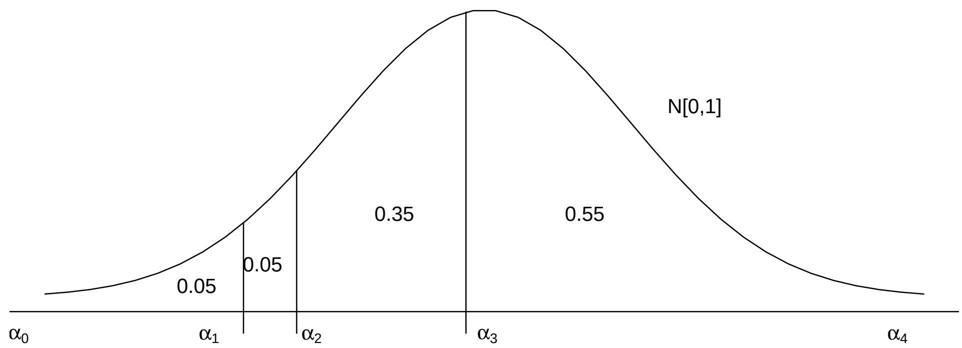

Figure 5.2: The underlying normal distribution with a threshold specification.

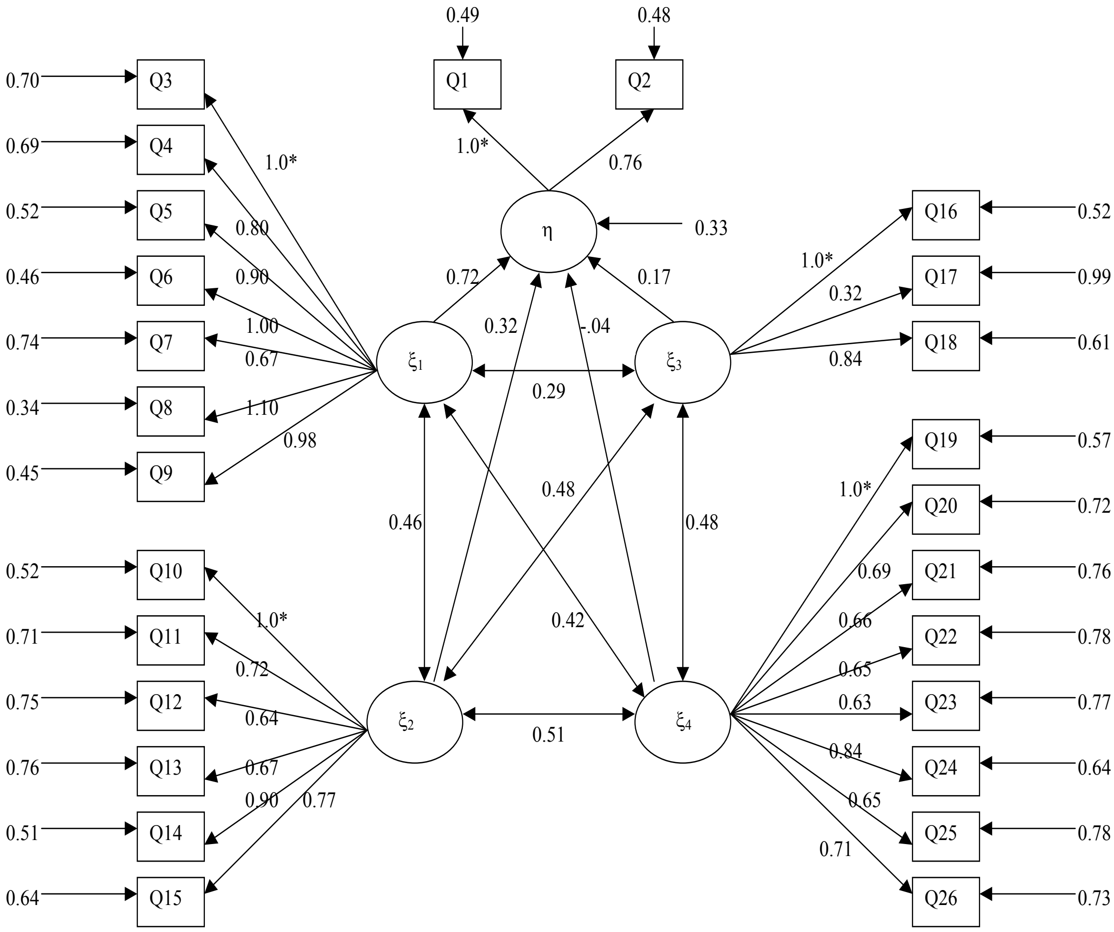

Figure 5.3: Path diagram and Bayesian estimates of parameters in the analysis of the QOL data. Note that Bayesian estimates of ϕ_{11}, ϕ_{22}, ϕ_{33}, and ϕ_{44} are 0.65, 0.71, 0.69, and 0.68, respectively. This figure is extracted from Lee (2007).