Chapter 5 Some Practical Structural Equation Models

In Chapter 2, we introduced

- linear SEMs
- nonlinear SEMs
- linear and nonlinear SEMs with fixed covariates

In Chapters 3 and 4, we introduced Bayesian methodologies, including

- Bayesian estimation
  - prior specification — informative and noninformative priors
  - posterior distribution
  - data agumentation and MCMC algorithm
- Model comparison statistics, including
  - Bayes factor
  - AIC, BIC, and DIC
  - L_ν-measure

In this chapter, we consider the following generalizations:

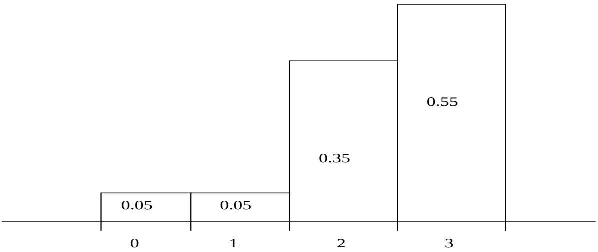
- SEMs with mixed continuous and ordered categorical variables
- SEMs with variables coming from an exponential family distribution
- SEMs with missing data

Due to the design of questionnaire and the nature of the problems on social, behavioral, and medical sciences, data are often ordered categorical. 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 them as continuous data from a normal distribution. This approach may not lead to a serious problem if the date are fairly symmetrical and with the highest frequencies at the center, which means most subjects choose the category 'no change'. 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.

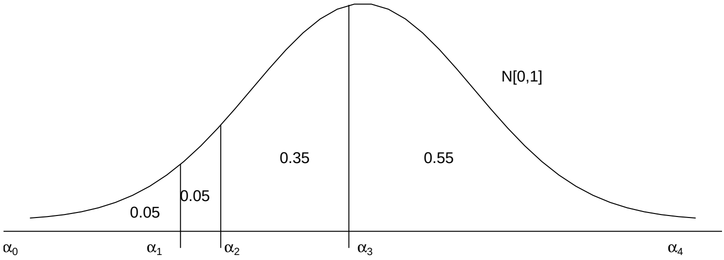
Suppose that for a given data set, the proportions of 0, 1, 2, 3 are 0.05, 0.05, 0.35, and 0.55, respectively. The discrete data are highly skewed to the left; see the histogram in Figure 5.1.



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 the following 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.

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

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

where $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\omega}_i$ and $\boldsymbol{\epsilon}_i$ are similar as before. Let $\boldsymbol{\omega} = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$, the structural equation is:

$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \tag{2}$$

where $\boldsymbol{\Pi}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\xi}_i$, and $\boldsymbol{\delta}_i$ are similar as before. Let $\boldsymbol{\Lambda}_\omega = (\boldsymbol{\Pi}, \boldsymbol{\Gamma})$, then Equation (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 whose continuous measurements are unobservable. The information of $\mathbf{y}$ is given by an observable ordered categorical vector $\mathbf{z} = (z_1, \cdots, z_s)^T$. 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:

$$\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_s \end{bmatrix} \quad \text{if} \quad \begin{matrix} \alpha_{1,z_1} \leq y_1 < \alpha_{1,z_1+1} \\ \vdots \\ \alpha_{s,z_s} \leq y_s < \alpha_{s,z_s+1} \end{matrix}, \tag{3}$$

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.

The SEM defined by (1) and (2) is not identified without imposing appropriate identification conditions. There are two kinds of indeterminacies involved in this model:

1. The 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.
2. The indeterminacy induced by the ordered categorical variables. To tackle this problem, we should note that
   (1) obtaining a necessary and sufficient condition for identification is difficult, so we aim to find a reasonable and convenient way to identify the model.
   (2) For an ordered categorical variable, the location and dispersion of its underlying continuous normal variable are unknown. It is common to take a unified scale to every ordered categorical variable.

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.

One method to solve the problem is to fix (μ, σ^2) at some constants. However, as these parameters are usually of main interest, we impose the identification conditions on the thresholds that are the less interesting nuisance parameters. That is, fixing 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. It can be applied to the multivariate case by imposing the above restrictions on the appropriate thresholds for every component in $\mathbf{z}$.

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.

Let

- $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ — observed continuous data matrix
- $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ — observed ordered categorical data matrix
- $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ — matrix of latent continuous measurement
- $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ — matrix of latent variables

In the Bayesian analysis, the observed data $[\mathbf{X}, \mathbf{Z}]$ are augmented with the latent data $[\mathbf{Y}, \boldsymbol{\Omega}]$ in the posterior analysis. The joint Bayesian estimates of $\boldsymbol{\Omega}$, 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 $\boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Lambda}_\omega, \boldsymbol{\Psi}_\epsilon$, and $\boldsymbol{\Psi}_\delta$ will be obtained.

In the Bayesian estimation, we draw samples from the posterior distribution $[\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \ \mathbf{Y}|\mathbf{X}, \mathbf{Z}]$ through the Gibbs sampler. At the jth iteration with current values $(\boldsymbol{\alpha}^{(j)}, \ \boldsymbol{\theta}^{(j)}, \ \boldsymbol{\Omega}^{(j)}, \ \mathbf{Y}^{(j)})$,

$$\begin{aligned} \text{Step (a)}: &\quad \text{Generate } \boldsymbol{\Omega}^{(j+1)} \text{ from } p(\boldsymbol{\Omega}|\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{X}, \mathbf{Z}); \\ \text{Step (b)}: &\quad \text{Generate } \boldsymbol{\theta}^{(j+1)} \text{ from } p(\boldsymbol{\theta}|\boldsymbol{\Omega}^{(j+1)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Y}^{(j)}, \mathbf{X}, \mathbf{Z}); \\ \text{Step (c)}: &\quad \text{Generate } (\boldsymbol{\alpha}^{(j+1)}, \mathbf{Y}^{(j+1)}) \text{ from } p(\boldsymbol{\alpha}, \mathbf{Y}|\boldsymbol{\theta}^{(j+1)}, \boldsymbol{\Omega}^{(j+1)}, \mathbf{X}, \mathbf{Z}). \end{aligned} \tag{4}$$

Convergence of the Gibbs sampler can be monitored by

1. the ‘estimated potential scale reduction (EPSR)’ values suggested by Gelman (1996);
2. plots of simulated sequences of the individual parameters with different starting points.

Based on the conjugate prior distributions of the parameters, conditional distributions required in the Gibbs sampler are presented in Appendix 5.1. Some of the conditional distributions are the familiar normal, Gamma, and inverted Wishart distributions. Drawing observations from them is straightforward and fast. However, $p(\boldsymbol{\alpha}, \mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\Omega})$ is nonstandard and complex. The MH algorithm 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 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 $[\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Y}|\mathbf{X}, \mathbf{Z}]$, the Bayesian infereces can be made accordingly.

Measures of quality of life (QOL) and health-related QOL are important for clinical work and the planning and evaluation of health care as well as for medical research. QOL is a multidimensional concept evaluated by a number of different latent constructs such as

- physical health status
- mental health status
- social relationships
- environmental conditions

As these latent constructs 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 to reflect the underlying latent construct of interest.

The WHOQOL-BREF instrument is a shorten version of WHOQOL-100 by selecting 24 ordered categorical items out of the 100 items for evaluating four latent constructs:

Q1-Q2 — overal QOL and health-related QOL

Q3-Q9 — physical health

Q10-Q15 — psychological health

Q16-Q18 — social relationship

Q19-Q26 — environment

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. For illustration, we only analyze a synthetic data set with sample size $n = 338$. Table 5.1 presents the frequencies of all the ordered categorical items, which shows that many items are skewed to the left.

		1	2	3	4	5
η	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
ξ_1	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
ξ_2	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
ξ_3	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
ξ_4	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

To illustrate the model comparison, we compare an SEM with four explanatory latent variables to another SEM with three explanatory latent variables.

Let M_1 be the SEM whose measurement equation is defined by

$$\mathbf{y}_i = \mathbf{\Lambda}_1 \boldsymbol{\omega}_{1i} + \boldsymbol{\epsilon}_i, \tag{5}$$

where $\boldsymbol{\omega}_{1i} = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4})^T$, $\boldsymbol{\epsilon}_i$ is distributed as $N[\mathbf{0}, \mathbf{\Psi}_{\epsilon 1}]$, and $\mathbf{\Lambda}_1^T$ is

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

ξ_3 ξ_4

The structural equation of M_1 is defined by

$$\eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \gamma_4 \xi_{i4} + \delta_i, \tag{6}$$

where $(\xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4})^T$ and δ_i are independently distributed as $N[\mathbf{0}, \mathbf{\Phi}_1]$ and $N[0, \sigma_{\delta 1}^2]$, respectively.

Let M_2 be the SEM whose measurement equation is defined by

$$\mathbf{y}_i = \mathbf{\Lambda}_2 \boldsymbol{\omega}_{2i} + \boldsymbol{\epsilon}_i, \tag{7}$$

where $\boldsymbol{\omega}_{2i} = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3})^T$, and $\boldsymbol{\epsilon}_i$ is distributed according to $N[\mathbf{0}, \mathbf{\Psi}_{\epsilon 2}]$. The first three columns of $\mathbf{\Lambda}_2$ are the same as those given in $\mathbf{\Lambda}_1$ 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

$$\eta_i = \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i3} + \delta_i, \tag{8}$$

where $(\xi_{i1}, \xi_{i2}, \xi_{i3})^T$ and δ_i are independently distributed as $N[\mathbf{0}, \mathbf{\Phi}_2]$ and $N[0, \sigma^2_{\delta 2}]$, respectively.

The Bayesian analysis is conducted using the conjugate prior distributions as follows:

- hyperparameters in the prior distributions of $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are all taken to be 0.8;
- those corrresponding to $\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ are $\{0.6, 0.6, 0.4, 0.4\}$;
- those corresponding to $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are $\rho_0 = 30$ and $\mathbf{R}_0^{-1} = 8\mathbf{I}$;
- $\alpha_{0\epsilon k} = \alpha_{0\delta k} = 10$, and $\beta_{0\epsilon k} = \beta_{0\delta k} = 8$;
- $\mathbf{H}_{0yk} = 0.25\mathbf{I}, \mathbf{H}_{0\omega k} = 0.25\mathbf{I}$, 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 M_t to directly link M_1 and M_2. Hence, we first compare M_1 with the following simple model M_0:

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

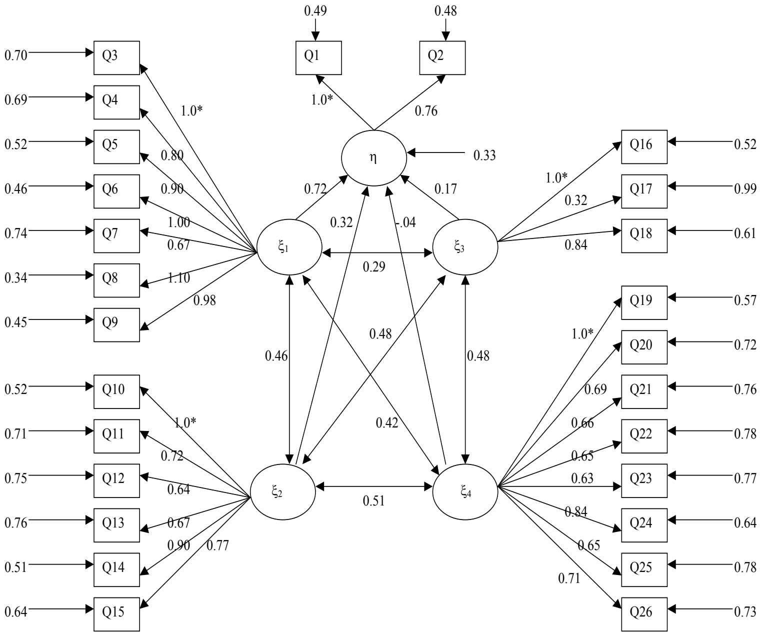
where $\boldsymbol{\epsilon}_i \stackrel{D}{=} N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$, and $\boldsymbol{\Psi}_\epsilon$ is a diagonal matrix. The link model to link M_0 and M_1 is defined by

$$\begin{aligned} M_t : \qquad & \mathbf{y}_i = t\boldsymbol{\Lambda}_1\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \\ & \eta_i = \gamma_1\xi_{i1} + \gamma_2\xi_{i2} + \gamma_3\xi_{i3} + \gamma_4\xi_{i4} + \delta_i. \end{aligned}$$

We obtained

- $\widehat{\log B_{10}} = 81.05$;
- $\widehat{\log B_{20}} = 57.65$ (through a comparison between M_2 and M_0);
- $\widehat{\log B_{12}} = \widehat{\log B_{10}} - \widehat{\log B_{20}}$ =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 the following Figure 5.3.


0.49
0.48
Q1
Q2
1.0*
0.76
η
0.33
Q3
Q4
Q5
Q6
Q7
Q8
Q9
0.70
0.69
0.52
0.46
0.74
0.34
0.45
1.0*
0.80
0.90
1.00
0.67
1.10
0.98
ξ1
0.72
0.32
-.04
0.17
ξ3
0.29
Q16
Q17
Q18
0.52
0.99
0.61
1.0*
0.32
0.84
0.46
0.48
0.42
0.48
Q10
Q11
Q12
Q13
Q14
Q15
0.52
0.71
0.75
0.76
0.51
0.64
1.0*
0.72
0.64
0.67
0.90
0.77
ξ2
0.51
ξ4
Q19
Q20
Q21
Q22
Q23
Q24
Q25
Q26
0.57
0.72
0.76
0.78
0.77
0.64
0.78
0.73
1.0*
0.69
0.66
0.65
0.63
0.84
0.65
0.71

Interpretations and conclusions:

1. All the factor loading estimates, except $\hat{\lambda}_{17,4}$ that associates with the indicator 'sexual activity', are high, indicating strong associations between each latent variable and their corresponding indicators.
2. 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.
3. The estimates of correlations $\{\phi_{12}, \phi_{13}, \phi_{14}, \phi_{23}, \phi_{24}, \phi_{34}\}$ among ξ_1 to ξ_4 are equal to $\{0.68, 0.43, 0.63, 0.69, 0.75, 0.70\}$, indicating that these explanatory latent variables are highly correlated.
4. The estimated structural equation

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

   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.

The software WinBUGS can produce Bayesian estimates of the structural parameters and latent variables for SEMs with ordered categorical variables. The WinBUGS code and the data related to the above QOL analysis will be given in a web-site housed in John-Wiley.

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.

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
- 'Success or Failure' in an experiment

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. Analysis of SEMs with dichotomous variables is similar to but not exactly the same as the analysis with ordered categorical variables.

Consider a common SEM defined by:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i,$$
$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i,$$

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 $k = 1, \cdots, p$,

$$z_{ik} = \begin{cases} 1, & \text{if } \ y_{ik} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The available observed data set is $\{\mathbf{z}_i, \ i = 1, \cdots, n\}$.

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. Then,

$$\begin{aligned}
&\Pr(z_{ik} = 1|\boldsymbol{\omega}_i, \mu_k, \mathbf{\Lambda}_k, \psi_{\epsilon k}) = \Pr(y_{ik} > 0|\boldsymbol{\omega}_i, \mu_k, \mathbf{\Lambda}_k, \psi_{\epsilon k}) \\
&= \Phi^*\{(\mathbf{\Lambda}_k^T/\psi_{\epsilon k}^{1/2})\boldsymbol{\omega}_i + \mu_k/\psi_{\epsilon k}^{1/2}\},
\end{aligned} \tag{10}$$

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 for any positive constant C,

$$C\mathbf{\Lambda}_k^T/(C\psi_{\epsilon k}^{1/2}) = \mathbf{\Lambda}_k^T/\psi_{\epsilon k}^{1/2} \quad \text{and} \quad C\mu_k/(C\psi_{\epsilon k}^{1/2}) = \mu_k/\psi_{\epsilon k}^{1/2}.$$

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. 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,

$$y_{ik} = \mu_k + \mathbf{\Lambda}_k \boldsymbol{\omega}_i + \epsilon_{ik} = \mu_k + u_{ik}.$$

Because

$$\begin{array}{llll} y_{ik} \geq 0 & \text{if and only if} & u_{ik} \geq -\mu_k; & \text{or equivalently} \\ z_{ik} = 1 & \text{if and only if} & u_{ik} \geq -\mu_k. & \end{array}$$

Consequently, $-\mu_k$ can be treated as the threshold corresponding to u_{ik}.

Note that 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

- $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ — the observed data set of dichotomous variables,
- $\boldsymbol{\theta}$ — the unknown parameter vector, which contains parameters in $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}_\omega$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Psi}_\delta$,
- $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ — the matrix of latent variables in the model,
- $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ — 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 $\boldsymbol{\Omega}$ and $\mathbf{Y}$; and a large sample of observations will be sampled from $p(\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Y}|\mathbf{Z})$ through the Gibbs sampler. At the jth iteration with $\boldsymbol{\theta}^{(j)}$, $\boldsymbol{\Omega}^{(j)}$, and $\mathbf{Y}^{(j)}$:

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

Conditional distributions in Step (a) and (b) can be obtained similarly as before. For $p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{Z})$, as $\mathbf{y}_i$ are mutually independent, we have

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

Moreover, it follows from the definition of the model that

$$[y_{ik}|\boldsymbol{\theta}, \boldsymbol{\omega}_i, \mathbf{z}_i] \stackrel{D}{=} \begin{cases} N[\mu_k + \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i, 1] I_{(-\infty,0]}(y_{ik}), & \text{if } z_{ik} = 0, \\ N[\mu_k + \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i, 1] I_{(0,\infty)}(y_{ik}), & \text{if } z_{ik} = 1, \end{cases} \tag{11}$$

where $I_A(y)$ is an indicator function.

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. Hence, it requires a large sample size to achieve accurate estimates.

A nonlinear SEM with fixed covariates on the basis of exponential family distributions (EFDs) is defined as follows. For $i = 1, \cdots, n$, let $\mathbf{y}_i = (y_{i1}, \cdots, y_{ip})^T$ be a vector of observed variables. 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.

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}:

$$\begin{aligned} p(y_{ik}|\boldsymbol{\omega}_i) &= \exp\{[y_{ik}\vartheta_{ik} - b(\vartheta_{ik})]/\psi_{\epsilon k} + c_k(y_{ik}, \psi_{\epsilon k})\}, \\ E(y_{ik}|\boldsymbol{\omega}_i) &= \dot{b}(\vartheta_{ik}), \quad \text{and} \quad \mathrm{Var}(y_{ik}|\boldsymbol{\omega}_i) = \psi_{\epsilon k}\ddot{b}(\vartheta_{ik}), \end{aligned} \tag{12}$$

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

- $\boldsymbol{\vartheta}_i = (\vartheta_{i1}, \cdots, \vartheta_{ip})^T$,
- $\mathbf{c}_{ik}$ $(m_k \times 1)$ — vectors of fixed covariates,
- $\mathbf{A}_k$ $(m_k \times 1)$ — vectors of unknown parameters,
- $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1^T, \cdots, \boldsymbol{\Lambda}_p^T)^T$ — a matrix of unknown parameters

We use the following model to relate ϑ_{ik} with $\mathbf{c}_{ik}$ and $\boldsymbol{\omega}_i$:

$$\vartheta_{ik} = \mathbf{A}_k^T \mathbf{c}_{ik} + \boldsymbol{\Lambda}_k^T \boldsymbol{\omega}_i. \tag{13}$$

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 $\mathbf{c}_{ik}$.

Let $\boldsymbol{\omega}_i = (\boldsymbol{\eta}_i^T, \boldsymbol{\xi}_i^T)^T$. To assess how the latent variables affect each other, we introduce the following nonlinear structural equation with fixed covariates:

$$\boldsymbol{\eta}_i = \mathbf{B}\mathbf{d}_i + \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i. \tag{14}$$

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 the structural eqution can be rewritten as

$$\boldsymbol{\eta}_i = \mathbf{\Lambda}_\omega \mathbf{G}(\boldsymbol{\omega}_i) + \boldsymbol{\delta}_i.$$

Note that the covariates in $\mathbf{d}_i$ may or may not be equal to those in $\mathbf{c}_{ik}$.

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: for $m = 0, \cdots, b$,

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

where $\{-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_b < \alpha_{b+1} = \infty\}$. This model can be identified by (i) fixing α_1 and α_b at preassigned values; and (ii) restricting the appropriate elements in $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\omega$ to fixed known values.

To implement the posterior analysis, proper conjugate prior distributions are taken for various unknown parameters. Let $\mathbf{\Psi}_\epsilon$ be the diagonal covariance matrix of the error measurements that correspond to the ordered categorical variables:

$$
\begin{aligned}
&\mathbf{A}_k \stackrel{D}{=} N[\mathbf{A}_{0k}, \mathbf{H}_{0k}],\ \psi_{\epsilon k}^{-1} \stackrel{D}{=} Gamma[\alpha_{0\epsilon k}, \beta_{0\epsilon k}],\\
&[\mathbf{\Lambda}_k|\psi_{\epsilon k}] \stackrel{D}{=} N[\mathbf{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}],\\
&\mathbf{\Phi}^{-1} \stackrel{D}{=} W_{q_2}[\mathbf{R}_0, \rho_0],\ \psi_{\delta k}^{-1} \stackrel{D}{=} Gamma[\alpha_{0\delta k}, \beta_{0\delta k}],\\
&[\mathbf{\Lambda}_{\omega k}|\psi_{\delta k}] \stackrel{D}{=} N[\mathbf{\Lambda}_{0\omega k}, \psi_{\delta k}\mathbf{H}_{0\omega k}].
\end{aligned}
\tag{15}
$$

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.

To handle the ordered categorical data, let

- ${\mathbf{y}_k^*}^T$ — the kth row of $\mathbf{Y}$ that is not directly observable,
- $\mathbf{z}_k$ — the observable ordered categorical vector,
- $\boldsymbol{\alpha}_k = (\alpha_{k,2}, \cdots, \alpha_{k,b_k-1})$.

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 $\alpha_{k,2} < \cdots < \alpha_{k,b_k-1}$. 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 $[\boldsymbol{\alpha}_k, \mathbf{y}_k^* | \mathbf{z}_k, \boldsymbol{\Omega}, \boldsymbol{\theta}]$ is derived in Appendix 5.2.

For model comparison, let

- $\mathbf{D}_o$ — the observed data, including the directly observable data and the ordered categorical data,
- $\mathbf{D}_m$ — the unobservable data, including latent variables and unobserved data that underlie the ordered categorical data.
- $\boldsymbol{\theta}^* = (\boldsymbol{\theta}, \boldsymbol{\alpha})$ — unknown parameter vector, where $\boldsymbol{\alpha}$ includes all unknown thresholds.

Suppose that $\mathbf{D}_o$ has arisen under one of the two competing models M_0 and M_1. For $k = 0, 1$, let $p(\mathbf{D}_o|M_k)$ be the probability density of $\mathbf{D}_o$ under M_k. The model comparison can be conducted using the Bayes factor

$$B_{10} = p(\mathbf{D}_o|M_1)/p(\mathbf{D}_o|M_0),$$

together with the path sampling procedure.

Let t be a continuous parameter in [0,1] to link the competing models M_0 and M_1, $p(\mathbf{D}_m, \mathbf{D}_o|\boldsymbol{\theta}^*, t)$ be the complete-data likelihood, and

$$U(\boldsymbol{\theta}^*, \mathbf{D}_m, \mathbf{D}_o, t) = d \log p(\mathbf{D}_m, \mathbf{D}_o|\boldsymbol{\theta}^*, t)/dt.$$

Let S be the number of fixed grids $\{t_{(s)}\}_{s=1}^S$ between [0,1] ordered as $0 = t_{(0)} < t_{(1)} < \cdots < t_{(S)} < t_{(S+1)} = 1$. $\log B_{10}$ can be computed as

$$\widehat{\log B_{10}} = \frac{1}{2}\sum_{s=0}^{S}(t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_{(s)}), \tag{16}$$

where

$$\bar{U}_{(s)} = J^{-1}\sum_{j=1}^{J} U(\boldsymbol{\theta}^{*(j)}, \mathbf{D}_m^{(j)}, \mathbf{D}_o, t_{(s)}),$$

in which $\{(\boldsymbol{\theta}^{*(j)}, \mathbf{D}_m^{(j)}),\ j = 1, \cdots, J\}$ are observations simulated from the joint conditional distribution $p(\boldsymbol{\theta}^*, \mathbf{D}_m|\mathbf{D}_o, t_{(s)})$.

In this simulation study, a data set $\mathbf{V} = \{\mathbf{v}_i,\ i = 1, \cdots, n\}$ was generated with $\mathbf{v}_i = (z_{i1}, z_{i2}, z_{i3}, y_{i4}, \cdots, y_{i9})^T$ as follows:

- For $k = 1,\ 2,\ 3,\ z_{ik}$ is dichotomous. unrankable
- For $k = 4, \cdots, 9,\ y_{ik}$ is binary with a distribution $B(1, p_{ik})$. That is,
  rankable

$$y_{ik} \propto \exp\{y_{ik}\vartheta_{ik} - \log(1 + e^{\vartheta_{ik}})\}$$

with

$$b(\vartheta_{ik}) = \log(1 + e^{\vartheta_{ik}}), \quad \text{and} \quad \vartheta_{ik} = \log(p_{ik}/(1 - p_{ik})).$$

Note that for each of these ~~binomial~~ binary. 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 corresponding to $(z_{i1}, z_{i2}, z_{i3})^T$. We assume that

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

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

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

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:

$$\eta_i = bd_i + \gamma_1 \xi_{i1} + \gamma_2 \xi_{i2} + \gamma_3 \xi_{i1} \xi_{i2} + \delta_i, \tag{17}$$

where

- d_i is assumed to come from $N[0, 1]$;
- 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;
- the true values of ϕ_{11}, ϕ_{21}, and ϕ_{22} in $\boldsymbol{\Phi}$ are 1.0, 0.3, and 1.0.

The following prior inputs in (15) are considered:

- elements in $\mathbf{A}_{0k}$, $\boldsymbol{\Lambda}_{0k}$, and $\boldsymbol{\Lambda}_{0\omega k}$ 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 $0.25\mathbf{I}$;
- $\alpha_{0\delta k} = 9$, $\beta_{0\delta k} = 3$, and $\rho_0 = 10$.

The purpose of this simulation study is to address the following questions:

(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.

Simulation settings:

- Three sample sizes $n = 200$, 800, and 2,000 were considered.
- For each sample size, the data were analyzed under 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;
  - (C) incorrectly treating all the binary variables as dichotomous variables.
- The results were obtained based on 100 replications.

R2WinBUGS is used to conduct the simulation study. 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

$\mathrm{M}(\hat{\theta}(r))$ — the mean of $\hat{\theta}(r)$;

$\mathrm{MC\text{-}SD}(\hat{\theta}(r))$ — the standard deviation of $\hat{\theta}(r)$;

$\mathrm{EST\text{-}SD}(\hat{\theta}(r))$ — the mean of the standard error estimates computed through $\sqrt{\widehat{\mathrm{Var}}(\boldsymbol{\theta}|\mathbf{V})}$.

The results are assessed through the following summary statistics:

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

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

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

Results corresponding to Case (A) and sample sizes of 200, 800, and 2,000 are reported in Table 5.2.

		$n = 200$			$n = 800$			$n = 2,000$		
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

Conclusions: in analyzing SEMs with binary or dichotomous variables:

(I) It requires comparatively larger sample sizes than the analysis with continuous variables.

(II) Except for situations with large sample sizes, the standard error estimates obtained from $\sqrt{\widehat{\mathrm{Var}}(\boldsymbol{\theta}|\mathbf{V})}$ overestimate the true standard deviations. Hence, 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 (not report here), which show that incorrectly treating binary data as dichotomous one, or vice versa would produce misleading results.

Missing data are very common in substantive research. Examples include:

A. Respondents in a household survey may refuse to report income.

B. Individuals in an opinion survey may refuse to express their attitudes toward some sensitive or embarrassing questions.

C. Patients in a follow-up treatment may miss some visits.

The main objectives of this section:

1. Define different types of missing data:
   - MAR (ignorable) — the probability of missingness depends on the fully observed data but not on the missing data themselves.
   - Nonignorable missing — the probability of missingness depends not only on the observed data but also on the missing data.
2. Introduce the Bayesian approach for analyzing SEMs with various kinds of missing data, including
   - SEMs with MAR missing data,
   - SEMs with nonignorable missing data.

Let

- $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$ — matrix of random vectors with $\mathbf{v}_i = (\mathbf{x}_i^T, \mathbf{y}_i^T)^T$;
- $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ and $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ — matrices of continuous variables whose exact measurements are observable and unobservable;
- $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)$ — observable ordered categorical data;
- $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$ — matrix of latent variables;
- $(\mathbf{X}_{obs}, \mathbf{X}_{mis})$ — observed and missing data sets of $\mathbf{X}$;
- $(\mathbf{Z}_{obs}, \mathbf{Z}_{mis})$ — observed and missing data sets of $\mathbf{Z}$;
- $(\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$ — observed and missing data sets of $\mathbf{Y}$ corresponding to $\mathbf{Z}_{obs}$ and $\mathbf{Z}_{mis}$.
- $\mathbf{V}_{obs} = \{\mathbf{X}_{obs}, \mathbf{Y}_{obs}\}$, $\mathbf{V}_{mis} = \{\mathbf{X}_{mis}, \mathbf{Y}_{mis}\}$;
- $\boldsymbol{\theta}^* = (\boldsymbol{\theta}, \boldsymbol{\alpha})$ — vectors of structural parameters and thresholds;
- $\{\mathbf{X}_{obs}, \mathbf{Z}_{obs}\}$ — observed data set;
- $\{\boldsymbol{\Omega}, \mathbf{X}_{mis}, \mathbf{Y}_{mis}, \mathbf{Y}_{obs}\} = \{\boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}\}$ — latent quantities.

What is the difference between unobservable & latent variable?

$$V = \begin{bmatrix} X_{obs} & Y_{obs} \\ X_{mis} & Y_{mis} \end{bmatrix} \leftarrow \Omega$$

Consider the Bayesian estimation by investigating the posterior distribution of $\boldsymbol{\theta}^*$ 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

- $p(\mathbf{X}_{obs}, \mathbf{Z}_{obs}|\boldsymbol{\theta}^*)$ — the observed-data likelihood;
- $p(\boldsymbol{\theta}^*)$ — the prior density of $\boldsymbol{\theta}^*$.

MCMC algorithm:

- Utilizing the data augmentation, $\{\mathbf{X}_{obs}, \mathbf{Z}_{obs}\}$ are augmented with the latent and missing quantities $\{\boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}\}$.
- A sufficiently large number of random observations will be simulated from $[\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}|\mathbf{X}_{obs}, \mathbf{Z}_{obs}]$. This task can be completed by a hybrid algorithm that combines the Gibbs sampler and the MH algorithm.

The following conditional distributions (see the derivation in the appendix) are required in the Gibbs sampler:

(A) $p(\boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs})$.

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

(C) $p(\mathbf{V}_{mis}|\boldsymbol{\theta}^*, \boldsymbol{\Omega}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs})$. Note that even $\mathbf{V}_{mis}$ may have many distinct missing patterns, its conditional distribution usually quite simple; see the following example.

(D) $p(\boldsymbol{\alpha}, \mathbf{Y}_{obs}|\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{V}_{mis}, \mathbf{X}_{obs}, \mathbf{Z}_{obs})$.

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 using the path sampling.

Note that the Bayesian method in handling MAR missing data requires only one additional step, sampling from $p(\mathbf{V}_{mis}|\theta^*, \mathbf{\Omega}, \mathbf{Y}_{obs}, \mathbf{X}_{obs}, \mathbf{Z}_{obs})$, in the Gibbs sampler.

Consider an SEM with mixed continuous and ordered categorical variables:

$$\mathbf{v}_i = \boldsymbol{\mu} + \mathbf{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i = \mathbf{\Pi}\boldsymbol{\eta}_i + \mathbf{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i.$$

Since $\mathbf{\Psi}_\epsilon$ is diagonal, $\mathbf{v}_{i,mis}$ is independent of $\mathbf{v}_{i,obs} = (\mathbf{x}_{i,obs}, \mathbf{y}_{i,obs})$, then

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

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

where p_i is the dimension of $\mathbf{v}_{i,mis}$, $\boldsymbol{\mu}_{i,mis}$ is a $p_i \times 1$ subvector of $\boldsymbol{\mu}$ with elements corresponding to observed components deleted, $\mathbf{\Lambda}_{i,mis}$ is a $p_i \times q$ submatrix of $\mathbf{\Lambda}$ with rows corresponding to observed components deleted, and $\mathbf{\Psi}_{i,mis}$ is a $p_i \times p_i$ submatrix of $\mathbf{\Psi}_\epsilon$ with the appropriate rows and columns deleted.

Data background and assumptions in the illustrative example:

- 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.
- 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.
- The missing values are assumed missing at random.

?

In the illustrative example,

- Six observed variables $(v_1, \cdots, v_6)$ are selected, where
    - v_1 and v_2 (ordered categorical) — 'worry about getting AIDS';
    - v_3 and v_4 (continuous) — 'aggressiveness';
    - v_5 and v_6 (ordered categorical) —'attitude to the risk of getting AIDS'.
- After deleting obvious outliers, the data set contains 1080 observations, only 754 of them are fully observed.
- The missing patterns (P) are complicated:

P	Sample size	Observed Variables 1	2	3	4	5	6	P	Sample size	Observed Variables 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	×	×

- To unify the scale, the continuous data are standardized.

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 f_{kh} are observed cumulative proportions of the categories with $z_k < h$.

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

$$\mathbf{v}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i,$$

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 are defined by the same measurement equation and the following different structural equations:

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

Bayes factors are estimated using the path sampling procedure with

- $S = 20$ and $J = 1,000$.
- A data-dependent prior inputs. We conduct an initial Bayesian estimation based on M_1 with noninformative priors to get some prior inputs. Then, we fixed $\mathbf{H}_{0yk} = \mathbf{I}$ and $\mathbf{H}_{0\omega k} = \mathbf{I}$; $\rho_0 = 10$, $\mathbf{R}_0^{-1} = 4\mathbf{I}$, $\alpha_{0\epsilon k} = \alpha_{0\delta k} = 8$, $\beta_{0\epsilon k} = \beta_{0\delta k} = 10$; and $\mathbf{\Lambda}_{0k}$ and $\mathbf{\Lambda}_{0\omega k}$ are taken as $\hat{\mathbf{\Lambda}}_k$ and $\hat{\mathbf{\Lambda}}_{\omega k}$ in the preliminary analysis.

To compare M_1 with the nonlinear models, say M_2, we use the link model M_t as follows:

$$M_t: \ \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + t\gamma_3\xi_1^2 + \delta.$$

Hence, when $t = 0$, $M_t = M_1$; and when $t = 1$, $M_t = M_2$. We obtain $\{\widehat{\log B_{21}}, \widehat{\log B_{31}}, \widehat{\log B_{41}}\} = \{2.303, 0.340, 0.780\}$, which suggests M_2 is the best model among $M_1, \cdots, M_4$.

To compare M_2 with more complex models, we consider:

$$\begin{aligned} M_5: & \quad \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1^2 + \gamma_4\xi_1\xi_2 + \delta, \\ M_6: & \quad \eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1^2 + \gamma_5\xi_2^2 + \delta. \end{aligned}$$

We obtain $\{\widehat{\log B_{52}}, \widehat{\log B_{62}}\} = \{0.406, 0.489\}$, which suggest selecting the simpler model M_2. The PP p-value corresponding to M_2 is 0.572, indicating that M_2 fits the data well.

The estimates of unknown parameters in M_2 are reported in Table 5.6.

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

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;
- overweight people are less likely to report their weights;
- respondents with significantly high or low salaries tend to not report the details.

Nonignorable missing data are more difficult to handle because we need to specify a missing data model to assess the influences of covariates on the probability of missingness for each individual. The most important methods for analyzing nonignorable missing data are developed by Ibrahim et al. (1999, 2001). These methods can be applied to handling nonignorable missing data in SEMs.

Let

- $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$ — observed data set, where $\mathbf{v}_i = (v_{i1}, \cdots, v_{ip})^T$;
- $\mathbf{r}_i = (r_{i1}, \cdots, r_{ip})^T$ — missing indicator vector such that

$$r_{ij} = \begin{cases} 1 & \text{if } \ v_{ij} \text{ is missing,} \\ 0 & \text{if } \ v_{ij} \text{ is observed.} \end{cases}$$

- $\mathbf{r} = (\mathbf{r}_1, \cdots, \mathbf{r}_n)$ — a matrix of indicators
- $(\mathbf{V}_{mis}, \mathbf{V}_{obs})$ — the missing and observed data.

For analyzing nonignorable missing data, the basic issues are

the miss proba dependent on V.

1. specifying a reasonable model for the conditional probability $p(\mathbf{r}|\mathbf{V}, \varphi)$, where φ is a vector of parameters involved in the missing data model;
2. 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 $p_{i1} + p_{i2} = p$.
- $\mathbf{V}_{obs} = \{\mathbf{v}_{i,obs}, i = 1, \cdots, n\}$, $\mathbf{V}_{mis} = \{\mathbf{v}_{i,mis}, i = 1, \cdots, n\}$.
- $[\mathbf{r}_i | \mathbf{v}_i, \boldsymbol{\omega}_i, \varphi]$ — the conditional distribution of $\mathbf{r}_i$ given $\mathbf{v}_i$ and $\boldsymbol{\omega}_i$ with a probability density function $p(\mathbf{r}_i | \mathbf{v}_i, \boldsymbol{\omega}_i, \varphi)$.
- $\boldsymbol{\theta}$ — 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 analysis of $p(\boldsymbol{\theta}, \varphi | \mathbf{V}_{obs}, \mathbf{r})$:

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

This 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 conditional distributions of the missing responses given the observed data, and inefficient sampling from those conditional distributions.

Since the observations are independent, we have

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

As $\boldsymbol{\Psi}_\epsilon$ is diagonal, when $\boldsymbol{\omega}_i$ is given, the components of $\mathbf{v}_i$ are independent. Hence, we use the following independent binomial model for the nonignorable missing mechanism:

$$p(\mathbf{r}|\mathbf{V}, \boldsymbol{\Omega}, \varphi) = \prod_{i=1}^{n}\prod_{j=1}^{p}\{\mathrm{pr}(r_{ij} = 1|\mathbf{v}_i, \boldsymbol{\omega}_i, \varphi)\}^{r_{ij}}\{1-\mathrm{pr}(r_{ij} = 1|\mathbf{v}_i, \boldsymbol{\omega}_i, \varphi)\}^{1-r_{ij}}. \tag{18}$$

Furthermore, the following logistic regression model is used:

$$\begin{aligned} m(\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}) &= \operatorname{logit}\{\operatorname{pr}(r_{ij} = 1 | \mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\} \\ &= \varphi_0 + \varphi_1 v_{i1} + \cdots + \varphi_p v_{ip} + \varphi_{p+1}\omega_{i1} + \cdots + \varphi_{p+q}\omega_{iq} = \boldsymbol{\varphi}^T \mathbf{e}_i, \end{aligned} \tag{19}$$

where $\mathbf{e}_i = (1, v_{i1}, \cdots, v_{ip}, \omega_{i1}, \cdots, \omega_{iq})^T$, and $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \cdots, \varphi_{p+q})^T$.

Sometimes it may be desirable to adopt a special case of (19):

$$m(\mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi}) = \operatorname{logit}\{\operatorname{pr}(r_{ij} = 1 | \mathbf{v}_i, \boldsymbol{\omega}_i, \boldsymbol{\varphi})\} = \varphi_0 + \varphi_1 v_{i1} + \cdots + \varphi_p v_{ip}. \tag{20}$$

Other missing mechanism model may be preferable for situations where one is certain about its specific form for the missing mechanism.

In the posterior analysis, the Gibbs sampler is used to generate observations from $[\boldsymbol{\Omega}, \mathbf{V}_{mis}, \boldsymbol{\theta}, \boldsymbol{\varphi} | \mathbf{V}_{obs}, \mathbf{r}]$. Specifically, we iteratively sample from the following conditional distributions:

- $p(\boldsymbol{\Omega}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r}) = p(\boldsymbol{\Omega}|\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r})$,
- $p(\mathbf{V}_{mis}|\ \mathbf{V}_{obs}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{r})$,
- $p(\boldsymbol{\varphi}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{r}) = p(\boldsymbol{\varphi}|\mathbf{V}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{r})$,
- $p(\boldsymbol{\theta}|\mathbf{V}_{obs}, \mathbf{V}_{mis}, \boldsymbol{\Omega}, \boldsymbol{\varphi}, \mathbf{r}) = p(\boldsymbol{\theta}|\mathbf{V}, \boldsymbol{\Omega})$.

Again, the Bayes factor is used for model comparison.

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

where

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

is the marginal density of M_k, and $p(\boldsymbol{\theta}_k, \boldsymbol{\varphi}_k)$ is the joint prior density.

For illustration, a small portion of the ICPSR data set collected by the World Values Survey 1981-1984 and 1990-1993 is analyzed. In this example,

- 8 variables are taken as $\mathbf{v} = (v_1, \cdots, v_8)$, where
    - v116-117 — 'job satisfaction',
    - v252-254 — 'job attitude',
    - v296,298,314 — 'morality (in relation to money)'.

  They are measured on a 10-point scale, so they are treated as continuous and standardized.
- The data corresponding to females in Russia, who either answered question 116 or 117, or both, are chosen. 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 have 69 different missing patterns. Considering that the questions are related to either personal attitudes or morality (in relation to money), the missing data are treated as nonignorable.

We use a nonlinear SEM with the following specifications to conduct the analysis. 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 variables can be interpreted as

- η — 'job satisfaction',
- ξ_1 — 'job attitude',
- ξ_2 — 'morality (in relation to money)'.

Consider 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:

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

$\varphi_1, \cdots, \varphi_8 = 0$ Not depend on

We obtained $\widehat{\log B^1_{ab}} = 47.34$ and $\widehat{\log B^1_{ac}} = 43.85$, which strongly support the missing data model M_a.

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 results of model comparison are reported in Table 5.7:

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

M_1, M_2, M_3 and M_4 are further compared under M_a, and M_2 is selected:

$$\widehat{\log B_{12}} = -1.29,\ \widehat{\log B_{32}} = -2.59,\ \widehat{\log B_{42}} = -1.51.$$

The Bayesian results in M_2 are presented in Table 5.8.

	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

Interpretations and conclusions:

- The estimates of φ_0, φ_2, φ_3, φ_4, φ_5, φ_6, and φ_8 are significantly different from zero, indicating the nonignorable missing data model is necessary.
- The factor loading estimates indicate strong associations between the latent variables and their indicators.
- $\hat{\phi}_{12} = 0.163$ 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 is similar to the interpretation of conventional nonlinear regression models.

In this chapter, we have discussed

- SEMs with mixed continuous and ordered categorical variables.
  Difficulty: simulate observations from $(\boldsymbol{\alpha}, \mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\Omega})$ efficiently using the MH algorithm.
- SEMs with variables coming from EFDs.
  Difficulty: most of full conditional distributions are nonstandard due to the nonlinear form of EFDs. Various MH algorithms are required.
- SEMs with ignorable and nonignorable missing data.
  Difficulties:
  1. Specify the missing data model(s).
  2. Derive the posterior distributions, some of which are nonstandard.
  3. Select missing data model(s).