Bayesian analysis of non-linear structural equation models with non-ignorable missing outcomes from reproductive dispersion models

Nian-Sheng Tang*, Xing Chen, Ying-Zi Fu
Department of Statistics, Yunnan University, Kunming 650091, China

ABSTRACT

Non-linear structural equation models are widely used to analyze the relationships among outcomes and latent variables in modern educational, medical, social and psychological studies. However, the existing theories and methods for analyzing non-linear structural equation models focus on the assumptions of outcomes from an exponential family, and hence can't be used to analyze non-exponential family outcomes. In this paper, a Bayesian method is developed to analyze non-linear structural equation models in which the manifest variables are from a reproductive dispersion model (RDM) and/or may be missing with non-ignorable missingness mechanism. The non-ignorable missingness mechanism is specified by a logistic regression model. A hybrid algorithm combining the Gibbs sampler and the Metropolis–Hastings algorithm is used to obtain the joint Bayesian estimates of structural parameters, latent variables and parameters in the logistic regression model, and a procedure calculating the Bayes factor for model comparison is given via path sampling. A goodness-of-fit statistic is proposed to assess the plausibility of the posited model. A simulation study and a real example are presented to illustrate the newly developed Bayesian methodologies.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

In modern educational, medical, social and psychological studies, various structural equation models have been developed to identify the latent variable from the manifest variables, and to assess the relationships of latent variables among themselves [1–5]. A lot of theories and methods have been proposed to analyze structural equation models in various fields on the basis of the assumptions of manifest variables from normal distribution. Recently, Lee & Tang [4] proposed a novel Bayesian method to analyze non-linear structural equation models with manifest variables from an exponential family. In particular, Lee, Song & Tang [5] introduced a Bayesian method to analyze a general structural equation model that accommodates the general non-linear terms of latent variables and covariates. Also, there are more than a dozen statistical software packages that have been developed to satisfy the strong demands in various fields, for example, EQS6 [1], LISREL [3], Mplus [6] and WinBUGS [7]. However, the above mentioned theories, methods and software packages can't be used to deal with structural equation models with manifest variables from a non-exponential family. Hence, it is important to develop a new approach to deal with more complex structural equation models.

In this paper, we consider non-linear structural equation models with manifest variables from a reproductive dispersion model [8] which includes a wide variety of distributions such as the normal, binomial, exponential, Poisson, Gamma, von...
Mises, simplex and Gumbel distributions. This family of distributions has been received a lot of attention in past decades. For example, Jorgensen [9,10,8] investigated asymptotic properties and saddlepoint approximation and applications of reproductive dispersion models; Tang et al. [11,12] proposed a non-linear reproductive dispersion model on the basis of the assumptions of responses from a reproductive dispersion model; Song [13] discussed application of reproductive dispersion models to longitudinal data. Although some theories and methods have been proposed to analyze non-linear structural equation models with exponential family distributions, procedures for analyzing non-linear structural dispersion models with manifest variables from reproductive dispersion models have not been developed. Hence, the main purpose of this paper is to develop a Bayesian method to obtain the joint Bayesian estimates of structural parameters and latent variables for non-linear structural equation models with manifest variables from reproductive dispersion models when some of manifest variables are missing.

As missing data are frequently encountered in behavioral, educational, medical, psychological, economical and social research, structural equation models with missing data have been received a lot of attention in recent years. For example, Lee & Song [14] presented a Monte Carlo EM (MCEM) procedure to analyze mixtures of structural equation models with ignorable missing data; Lee, Song & Lee [15] extended ML estimate and model comparison procedure of Lee & Song [14] to non-linear structural equation models with ignorable missing data; Lee & Song [16] gave a Bayesian model comparison procedure for non-linear structural equation models with missing continuous and ordinal categorical data. The above mentioned works were developed under the missing at random (MAR) assumption with an ignorable missing mechanism as defined in Little & Rubin [17]. However, in medical, economical, educational, and behavioral studies, missing data are often non-ignorable in the sense that the reason for missingness often depends on missing value themselves [18]. Recently, Lee & Tang [18] proposed a hybrid Bayes procedure to analyze non-linear structural equation models with non-ignorable missing data by combining the Gibbs sampler and the Metropolis–Hastings algorithm on the basis of a non-ignorable missingness mechanism which is specified by a logistic regression model; Lee & Tang [19] developed a Bayesian approach for non-linear structural equation models with covariates and mixed continuous and ordered categorical outcomes in the presence of missing observations and missing covariates that are missing with a non-ignorable mechanism on the basis of MCMC algorithm. However, there is no work done on non-linear structural equation models with manifest variables from reproductive dispersion models in the presence of missing observations that are missing with a non-ignorable mechanism.

In this paper, we will develop a Bayesian procedure to investigate the estimation of the models and model comparison.

The rest of the paper is organized as follows: In Section 2, we introduce a non-linear structural equation model with manifest variables from a reproductive dispersion model that can be missing with a non-ignorable mechanism. The specification of the non-ignorable missingness mechanism models is also discussed in Section 2. In Section 3, a Metropolis–Hastings-within-Gibbs algorithm is developed for estimation and model comparison, novel conditional distributions for Gibbs sampler and Metropolis–Hastings algorithm are also given in Section 3. A partial posterior predictive p-value [20,4] is proposed to assess the plausibility of the posited model, and a path sampling [21] procedure is proposed to calculate the Bayes factor for model comparison in Section 3. Numerical illustrations, which include a simulation study and a real example, are presented in Section 4. Some concluding remarks are given in Section 5.

2. Model and notation

Let \( y_i = (y_{i1}, \ldots, y_{ip})^T \) be a \( p \times 1 \) vector of manifest variables measured on the \( i \)th individual for \( i = 1, \ldots, n \), and let \( \xi_i \) be a \( q \times 1 \) vector of latent variables corresponding to \( y_i \). The main purpose of this paper is to identify the relationship between the manifest variables in \( y_i \) and the latent variables in \( \xi_i \). Given latent variables in \( \xi_i \), we assume that \( y_{i1}, \ldots, y_{ip} \) are conditionally independent, and each \( y_{ik} (k = 1, \ldots, p) \) is distributed as a reproductive dispersion model with parameters \( \psi_k \) and \( \mu_{ik} \), which is a function of the vector of latent variables \( \xi_i \). That is, for \( k = 1, \ldots, p \) and \( i = 1, \ldots, n \), \( y_{ik} \) has the following probability density function

\[
p(y_{ik} | \psi_k) = a(y_{ik}; \psi_k) \exp \left\{ -\frac{1}{2\psi_k} d(y_{ik}; \mu_{ik}) \right\}.
\] (1)

For the sake of simplicity, let \( c(y_{ik}; \psi_k) = \log a(y_{ik}; \psi_k) \), thus the Eq. (1) can be rewritten as

\[
p(y_{ik} | \psi_k) = \exp \left\{ -\frac{1}{2\psi_k} d(y_{ik}; \mu_{ik}) \right\}.
\] (2)

where \( \mu_{ik} \) is the location parameter and may represent the mean of the distribution; \( \psi_k \in \Lambda (\Lambda \subset \mathbb{R}^+) \) is usually referred to as the dispersion parameter which is known or can be estimated separately; \( a(\cdot) > 0 \) is a suitable known function; \( d(y; \mu) \) is a unit deviance defined on \( C \times \Omega \) (here \( \Omega \subseteq C \subseteq \mathbb{R} \) is an open interval, and \( C \) is also an interval), and satisfies \( d(y; y) = 0 \forall y \in \Omega \) and \( d(y; \mu) > 0 \forall y \neq \mu \), and is twice continuously differentiable with respect to \( y, \mu \) on \( C \times \Omega \). Eq. (1) includes normal distribution, extreme value distribution and exponential family distribution as its special case [8,11,12].

Similar to [22,18], we model the relationship between \( \mu_i = (\mu_{i1}, \ldots, \mu_{ip})^T \) and \( \xi_i \) via the following measurement equation:

\[
\mu_i = (\mu_{i1}, \ldots, \mu_{ip}) = u + \Lambda \xi_i,
\] (3)
where \( u \) is a \( p \times 1 \) unknown parameter vector, \( A \) is a \( p \times q \) unknown factor loading matrix. Let \( \xi_i = (\xi_{i(1)}, \xi_{i(2)}) \) be a partition of \( \xi_i \) into endogenous latent variables in \( \xi_{i(1)} (q_1 \times 1) \) and exogenous variables in \( \xi_{i(2)} (q_2 \times 1) \), where \( q_1 + q_2 = q \). Thus, the relationship between \( \xi_{i(1)} \) and \( \xi_{i(2)} \) can be modeled via the following non-linear structural equation:

\[
\xi_{i(1)} = \Pi \xi_{i(1)} + \Gamma h(\xi_{i(2)}) + \delta_i,
\]

where \( h(\xi_{i(2)}) = (h_1(\xi_{i(2)}), \ldots, h_l(\xi_{i(2)}))' \) is a \( t \times 1 \) \((t \geq q_2)\) vector-valued function containing non-zero differentiable functions \( h_1, \ldots, h_l, \Pi (q_1 \times q_1) \) and \( \Gamma (q_1 \times t) \) are matrices of unknown regression coefficients of \( \xi_{i(1)} \) on \( \xi_{i(1)} \) and \( h(\xi_{i(2)}) \); \( \delta_i \) is a vector of unknown parameters.

It is assumed that \( \xi_{i(2)} \) and \( \delta_i \) are independently distributed as \( N(0, \Phi) \) and \( N(0, \Psi_\delta) \), respectively, where \( \Psi_\delta \) is a diagonal matrix and \( \Phi \) is a positive definite matrix. Let \( A_\xi = (\Pi, \Gamma) \) and \( g(\xi_i) = (\xi_{i(1)}', h(\xi_{i(2)}))' \), then the non-linear structural equation given in (4) can be rewritten as:

\[
\xi_{i(1)} = A_\xi g(\xi_i) + \delta_i.
\]

It is easily seen from Eqs. (1), (3) and (4) that the above defined non-linear structural equation model reduces to the model discussed by Lee & Zhu [23] if \( d(y_h; \mu_{ik}) = (y_h - \mu_{ik})^2 \); and it also reduces to the model discussed by Lee & Tang [4] if \( d(y_h; \mu_{ik}) = -2(y_h\mu_{ik} - b(\mu_{ik})) \) and \( h(\xi_{i(2)}) = \xi_{i(2)} \) where \( b(\mu_{ik}) \) is some specific differentiable function. Therefore, the above introduced model is an extension of non-linear structural equation models with manifest variables from a normal distribution family [23] and structural equation models with manifest variables from an exponential distribution family [4].

In this paper, we consider the situation where the manifest vector \( y_i \) is incompletely observed with a non-ignorable mechanism. Let \( y_i = (y_{im}, y_{in})' \), where \( y_{im} \) is a \( p_1 \times 1 \) vector of observed manifest variable, \( y_{in} \) is a \( p_2 \times 1 \) vector of missing components of the manifest vector \( y_i \), and \( p_1 + p_2 = p \). Here, we assume an arbitrary pattern of missing data in \( y_i \), and thus \( y_i = (y_{im}, y_{in})' \) may represent some permutation of the indices of the original \( y_i \). Let \( r_i = (r_{i1}, \ldots, r_{ip}) \) be a vector of missing indicators for \( y_i \) such that \( r_{ij} = 1 \) if \( y_{ij} \) is missing and \( r_{ij} = 0 \) if \( y_{ij} \) is observed. Let \( (r_i \mid y_i, \xi_i, \phi) \) be the conditional distribution of \( r_i \) given \( y_i \) and \( \xi_i \), where \( \phi \) is an unknown parameter vector in conditional probability density function \( p(r_i \mid y_i, \xi_i, \phi) \). The missing data mechanism is decided by this conditional distribution. Let \( \theta \) be the structural parameter vector that contains all unknown distinct parameters in \( u, A, \Psi = (\psi_1, \ldots, \psi_p)' \), \( A_\xi, \Psi_\phi \) and \( \Phi \).

The main purpose of this paper is to present a Bayesian approach to analyze the above defined model on the basis of the missing data indicator \( r = (r_{i1}, \ldots, r_{ip}) \) and the observed data \( Y_o = \{y_{o1}, \ldots, y_{on}\} \). To obtain Bayesian estimates of unknown parameters \( \theta \) and \( \phi \) and to make inference on the above defined model, we need to sample observations from the posterior distributions of \( \theta \) and \( \phi \). According to the definition of the model, the joint posterior density of parameters \( \theta \) and \( \phi \) on the basis of observed data \( Y_o \) and \( r \) is given by

\[
p(\theta, \phi \mid Y_o, r) \propto p(Y_o, r \mid \theta, \phi)p(\theta, \phi),
\]

where \( p(\theta, \phi) \) denotes the joint prior distribution of \( \theta \) and \( \phi \). It is rather difficult to obtain a closed form of integral (5) because of the complexity of the reproductive dispersion model (1), and non-linear relationship of structural equation model (4) and missingness data mechanism involved. Clearly, Eq. (5) involves specification of non-ignorable missingness mechanism model \( p(r_i \mid y_i, \xi_i, \phi) \). In general, we can consider any general model for \( p(r_i \mid y_i, \xi_i, \phi) \). But, a too complicated or large model may result in unidentification of the model and may also induce difficulty in deriving the corresponding conditional distribution of the manifest given the observed data and/or inefficient sampling from that conditional distribution [18]. According to assumption of the model, we know that given \( \xi_i \), the components of \( y_i \) are conditionally independent which motivates us to assume that for \( j \neq l \), the conditional distributions of \( r_{ij} \) and \( r_{il} \) given \( \xi_i \) are independent. Then we have

\[
p(r \mid Y, F, \xi, \phi) = n \prod_{i=1}^{n} p(r_i \mid y_i, \xi_i, \phi) = \prod_{i=1}^{n} \prod_{j=1}^{p} p(r_{ij} \mid y_i, \xi_i, \phi),
\]

where \( Y = (y_{i1}, \ldots, y_{in})' \) and \( F = (\xi_1, \ldots, \xi_n) \). Following Ibrahim, Chen & Lipsitz [24], we consider the following non-ignorable missingness mechanism

\[
p(r_{ij} \mid y_i, \xi_i, \phi) = \{pr(r_{ij} = 1 \mid y_i, \xi_i, \phi)\}^{r_{ij}}(1 - pr(r_{ij} = 1 \mid y_i, \xi_i, \phi))^{1-r_{ij}},
\]

where \( pr(r_{ij} = 1 \mid y_i, \xi_i, \phi) \) can be formulated by the following logistic regression model

\[
m(y_i, \xi_i, \phi) = \logit[pr(r_{ij} = 1 \mid y_i, \xi_i, \phi)]
\]

\[
\triangleq \phi_0 + \phi_1 y_{i1} + \cdots + \phi_p y_{ip} + \phi_{p+1} \xi_{i1} + \cdots + \phi_{p+q} \xi_{iq}
\]

\[
\triangleq \phi_0 + \phi_{\xi} \xi_i + \phi_{\psi} \psi_i + \phi_{\f} \f_i,
\]
where $\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_{p+q})^T$, $\varphi_m = (\varphi_{p+1}, \ldots, \varphi_{p+q})^T$, $\varphi_o$ and $\varphi_{ml}$ are vectors corresponding to $y_{oi}$ and $y_{mi}$, respectively; and $\omega_i = (1, y_{i1}, \ldots, y_{ip}, \xi_{i1}, \ldots, \xi_{iq})^T$. As pointed out by Lee & Tang [18], we could relax the above assumption on the independence of $r_i$ and $r_m$ in practical application by specifying the following non-ignorable missingness mechanism

$$
p(r|r, F, \varphi) = \prod_{i=1}^{n} p(r_{i1}, \ldots, r_{ip}|y_i, \xi_i, \varphi) = \prod_{i=1}^{n} p(r_{i1}|y_i, \xi_i, \varphi_1)p(r_{i2}|r_{i1}, y_i, \xi_i, \varphi_2) \cdots p(r_{ip}|r_{i1}, \ldots, r_{i,p-1}, y_i, \xi_i, \varphi_p),
$$

where $\varphi = (\varphi_1^T, \ldots, \varphi_p^T)^T$. In this case, we may use a logistic regression model similar to Eqs. (6) and (7) to formulate $p(r_j|r_{11}, \ldots, r_{i,j-1}, y_i, \xi_i, \varphi_j)$.

### 3. Bayesian analysis of the model

Let $\lambda_k^T$ and $\lambda_{2k}^T$ be the $k$th row vectors of matrices $A$ and $A_\delta$, respectively; and let $\psi_{ik}$ be the $k$th diagonal elements of matrix $\Psi_i$. Let $Y_m = \{y_{m1}, y_{m2}, \ldots, y_{mn}\}$ be set of missing data corresponding to $Y$. To develop an effective Bayesian procedure to analyze the above introduced model and to reduce computational burden, we augment the observed data $Y_o$ and the missing data indicator $r$ with the missing quantities $\{Y_m, F\}$ to produce a complete-data set $\{Y, F, r\}$. Thus, the joint posterior distribution of interest is $p(Y_m, F, \theta, \varphi|Y_o, r)$. Clearly, the posterior density $p(Y_m, F, \theta, \varphi|Y_o, r)$ is easier to handle than $p(\theta, \varphi|Y_o, r)$. However, it is still rather difficult to sample observations from $p(Y_m, F, \theta, \varphi|Y_o, r)$ because of complexity of the considered model. Hence, the Gibbs sampler [25] is adopted to sample a sequence of random observations from the joint posterior density $p(Y_m, F, \theta, \varphi|Y_o, r)$, and the Bayesian estimates, the partial posterior predictive $p$-value and Bayes factor for model comparison can be obtained from the sampled observations. In this algorithm, observations $\{F, Y_m, \theta, \varphi\}$ are iteratively generated from the following conditional distributions: $p(F|Y_m, Y_o, r, \theta, \varphi) = p(F|Y_m, r, \theta, \varphi)$, $p(Y_m|F, Y_o, r, \theta, \varphi)$, $p(\theta|Y_m, Y_o, F, r, \varphi)$, $p(\varphi|Y_m, Y_o, F, r, \theta) = p(\varphi|Y_o, F, r)$ because $\theta$ has no relationship with the non-ignorable missingness mechanism model, and $p(\varphi|Y_m, Y_o, F, r, \theta) = p(\varphi|Y_o, F, r)$. Based on our experience, the Bayesian results do not depend on the order of $\{F, Y_m, \theta, \varphi\}$ in simulating a sequence of random observations. Geman & Geman [25] pointed out that under some regularity conditions and for a sufficiently large $J$, the sequence of observations $\{(F^{(j)}, Y_m^{(j)}, \theta^{(j)}, \varphi^{(j)}): j = 1, \ldots, J\}$ can be regarded as a sample generated from the joint posterior density $p(F, Y_m, \theta, \varphi|Y_o, r)$. In practice application, convergence of the algorithm can be monitored by the “estimated potential scale reduction (EPSR)” values [26]. Convergence is attained if all the EPSR values are less than 1.2. Also, convergence of the algorithm can be monitored by inspecting plots of the simulated sequences from different starting values. In this case, the algorithm is claimed to be convergence if the simulated sequences mix well.

#### 3.1. Conditional distributions

Conditional distributions required in the implementation of the Gibbs sampler will be briefly introduced on the basis of non-ignorable missingness mechanism (7) in this subsection. First, we consider conditional distribution $p(F|Y, r, \theta, \varphi)$. Note that given $\xi_i, y_i$ are conditionally independent, and $\hat{\xi}_i$ are also conditionally independent. Then, we have

$$
p(F|Y, r, \theta, \varphi) = \prod_{i=1}^{n} p(F|y_i, \xi_i, \theta, \varphi) \propto \prod_{i=1}^{n} p(y_i|\xi_i, \theta) p(\xi_{i1}|\xi_{i2}) p(\xi_{i2}|\theta) p(r_i|y_i, \xi_i, \varphi).
$$

Then, $p(F|y_i, r_i, \theta, \varphi)$ is proportional to

$$
\exp \left\{ -\frac{1}{2} \xi_{i2}^T \Lambda_\xi^{-1} \xi_{i1} - \frac{1}{2} \xi_{i1}^T \Lambda_\xi^{-1} \xi_{i1} - A_\delta g(\xi_i) \right\} + \sum_{k=1}^{p} \left[ -\frac{1}{2} \psi_k \left( d(y_{ik}; \mu_k) + c(y_{ik}; \psi_k) \right) + \left( \frac{1}{2} \sum_{k=1}^{p} \sigma_{ik} \right) \varphi^T \omega_i - p \log(1 + \exp(\varphi^T \omega_i)) \right],
$$

where $\mu_k$ is the $k$th component of $\mu = u + A_\xi \xi_i$.

Next, we consider conditional distribution $p(Y_m|Y_o, F, r, \theta, \varphi)$. As $\xi_i$ is given, $y_{i1}, \ldots, y_{in}$ are conditionally independent, thus $y_{m1}, \ldots, y_{mn}$ are also conditionally independent. More important, when $\xi_i$ is given, $y_{mi}$ is independent of $y_{oi}$ for $i = 1, \ldots, n$. Then, we have

$$
p(Y_m|Y_o, r, F, \theta, \varphi) = \prod_{i=1}^{n} p(y_{mi}|y_{oi}, r_i, \xi_i, \theta, \varphi) \propto \prod_{i=1}^{n} p(y_{mi}|\xi_i, \theta) p(r_i|y_i, \xi_i, \varphi).
$$
According to the definitions of models for \( y_{mi} \) and \( r_i \), it follows from Eqs. (2), (6) and (7) that

\[
p(y_{mi}|y_{ui}, r_i, \xi, \theta, \varphi) \propto \exp \left\{ \sum_{j=1}^{n_i} \left( -\frac{1}{2} \psi_{kj} d(y_{ik}, \mu_{ik} \xi_j) + c(y_{ik}, \psi_{kj}) \right) + \left( \sum_{k=1}^{p} r_{ik} \right) \varphi^T \omega_i - p \log(1 + \exp(\varphi^T \omega_i)) \right\},
\]

where \( h_i = p z_i \), \( \mu_{ik} \) is the \( k \)th component of vector \( \mu_i = u + \Lambda \xi_j \) corresponding to missing element \( y_{ik}, y_{ik} \) is the \( k \)th element of \( y_i, \psi_{kj} \) is the \( k \)th diagonal element of \( \Psi, k_1 < k_2 < \cdots < k_p \) are the indices of \( y_{mi} \) corresponding to missing components of \( y_i \).

The conditional distribution \( p(\theta|Y, F) \) of \( \theta \) given \( Y \) and \( F \) depends on the prior distribution \( p(\theta) = (\theta^T, \theta^T)^T \) in which \( \theta_1 \) contains all unknown distinct parameters in \( u, \Psi, \Lambda, \) and \( \theta_2 \) contains all unknown distinct parameters in \( \Psi, \Lambda \) and \( \Phi \). To derive the conditional distribution of \( p(\theta|Y, F) \), following Lindsley & Smith [27] and other Bayesian analyses of non-linear structural equation models, we consider the following prior distributions for \( u, \Psi, \) and \( \Lambda \):

\[
p(u) \overset{\text{D}}{=} N(u_0, \Sigma_0), \quad p(\varphi)|\psi_k \overset{\text{D}}{=} \text{Gamma}[\alpha_{0\psi k}, \beta_{0\psi k}], \quad p(\lambda_k|\psi_k) \overset{\text{D}}{=} N[\lambda_{0k}, \psi_k H_{0\psi k}],
\]

where \( u_0, \alpha_{0\psi k}, \beta_{0\psi k}, \lambda_{0k} \) and positive definite matrices \( \Sigma_0 \) and \( H_{0\psi k} \) are hyperparameters whose values are assumed to be given by the prior information. As in [23], we assume that \( \lambda_k \) is independent of \( \lambda_l \) for \( k \neq l \). It is easily shown from the above assumptions that the conditional distributions \( p(u|Y, F, \Lambda, \Psi), p(\psi_k|Y, F, \Lambda, u) \) and \( p(\lambda_k|Y, F, \Psi, u) \) are given by:

\[
p(u|Y, F, \Lambda, \Psi) \propto \exp \left\{ -\frac{1}{2} \sum_{k=1}^{n} \sum_{i=1}^{p} d(y_{ik}; \mu_{ik}) - \frac{1}{2} (u - u_0)^T \Sigma_0^{-1} (u - u_0) \right\},
\]

\[
p(\psi_k^{-1}|Y, F, \Lambda, u) \propto (\psi_k^{-1})^{\alpha_{0\psi k} - 1} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} d(y_{ik}; \mu_{ik}) + \frac{1}{2} \sum_{i=1}^{n} c(y_{ik}; \psi_{kj}) - \beta_{0\psi k}/\psi_k \right\},
\]

\[
p(\lambda_k|Y, F, \Psi, u) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} d(y_{ik}; \mu_{ik}) + (\lambda_k - \lambda_{0k})^T H_{0\psi k}^{-1} (\lambda_k - \lambda_{0k}) \right\}
\]

with \( \mu_{ik} = u_k + \lambda_1^T \xi_i. \) In particular, if \( y_{ik}|\xi_i \) is distributed as simplex distribution with parameters \( \mu_{ik} \) and \( \psi_k \), then we have

\[
p(\psi_k^{-1}|Y, F, \Lambda, u) \sim \text{Gamma} \left[ \alpha_{0\psi k} + n/2, \beta_{0\psi k} + \sum_{i=1}^{n} d(y_{ik}; \mu_{ik})/2 \right].
\]

Similarly, to derive the conditional distribution \( p(\theta_2|Y, F) \), we consider the following conjugate type prior distributions for \( \varphi, \Lambda_\xi, \alpha_\xi \) and \( \Phi \):

\[
p(\varphi)|\psi_k \overset{\text{D}}{=} \text{Gamma}[\alpha_{0\phi}, \beta_{0\phi}], \quad p(\lambda_\xi|\psi_k) \overset{\text{D}}{=} N[\lambda_{0\xi}, \psi_k H_{0\phi}], \quad p(\Phi) \overset{\text{D}}{=} \text{IW}(R_0, q_0, q_2),
\]

where \( \alpha_{0\phi}, \beta_{0\phi}, \lambda_{0\phi}, \rho_0 \) and positive definite matrices \( H_{0\phi} \) and \( R_0 \) are hyper-parameters whose values are assumed to be given by the prior information; and \( \text{IW}[-1, \cdot] \) denotes inverted Wishart distribution. These conjugate type prior distributions are flexible, and for situations with a reasonable amount of data available, the hyper-parameter values scarcely affect the analysis. For the sake of simplicity, we assume that for \( h \neq k, (\psi_{sk}, \lambda_{sk}) \) are independent of \( (\psi_{sk}, \lambda_{sk}) \). Then, it follows from the above prior distributions that

\[
p(\psi_{sk}^{-1}|F) \sim \text{Gamma} \left[ \frac{n}{2} + \alpha_{0\phi}, \beta_{0\phi} \right], \quad p(\lambda_{sk}|\psi_k) \overset{\text{D}}{=} N[\psi_{sk}, \psi_k \Omega_{sk}], \quad p(\Phi|F) \sim \text{IW}(F_k \Sigma_k^T + R_0, n + p_0, q_2),
\]

where \( \Omega_{sk} = (H_{0sk}^{-1} + g(\xi_j)g(\xi_j)^T)^{-1}, \psi_{sk} = \Omega_{sk} \Omega_{sk}^{-1} + g(\xi_j) f_{sk}, \beta_k = \beta_{0k} + (f_{sk} f_{sk}^T - v_{sk}^T \Omega_{sk}^{-1} v_{sk} + \lambda_{0sk} H_{0sk}^{-1} \lambda_{0sk})/2, f_{sk} \) is the kth row vector of \( F_k = (\xi_{1h1}, \ldots, \xi_{nh1}) \) which is submatrix of \( F \).

Finally, we consider the conditional distribution of \( \varphi \) given \( Y, F, \theta \) and \( r \). Let \( p(\varphi) \) be the prior density of \( \varphi \) with distribution \( N(\varphi^0, V) \) [19], where \( \varphi^0 \) and \( V \) are the hyper-parameters whose values are assumed to be given by the prior information. Since the distribution of \( r \) only involves \( Y, F \) and \( \varphi, \) it is reasonable to assume that the prior distribution of \( \varphi \) is independent of the prior distribution of \( \theta. \) Under the above assumptions, we have

\[
p(\varphi|Y, r, \theta) \propto p(r|Y, F, \varphi, \theta) p(\varphi).
\]

Based on the above prior distribution and the fact that the distribution of \( r_{ij} \) only involves \( y_{ij}, \xi_j \) and \( \varphi \) for \( j \in \{1, \ldots, p\}, \) it follows from the independence of \( r_{ij} \) and \( r_{ih} \) for any \( j \neq h \) that

\[
p(\varphi|Y, r) \propto \prod_{i=1}^{n} \exp \left\{ \left( \sum_{j=1}^{p} \sum_{k=1}^{r_{ik}} \varphi^T \omega_i - \frac{1}{2} (\varphi - \varphi^0)^T V^{-1} (\varphi - \varphi^0) \right) \right\}
\]

\[
\prod_{i=1}^{n} (1 + \exp(\varphi^T \omega_i))^p.
\]
Here, conditional distributions of $\Lambda$ and $\Omega_{\xi}$ are presented for the case without fixed parameters. In practical application, some elements in $\Lambda$ and $\Omega_{\xi}$ may be fixed values. In this case, the approach presented in Lee & Zhu [28] can be employed to solve their conditional distributions. Also, the conditional distributions associated with non-informative prior distributions can be obtained by taking huge variances in the corresponding prior distributions.

3.2. Implementation

It is easily seen from (14) and (15) that the conditional distributions of $p(\psi_{k}^{-1}F)$, $p(\lambda_{k}\text{exp}(F, \psi_{k}))$ and $p(\Phi|F_{2})$ are the familiar distributions. In particular, if $y_{u1}\xi_{i}$ is distributed as a simplex distribution, then it is easily seen from (13) that the conditional distribution of $p(\psi_{k}^{-1}Y, F, \Lambda, u)$ is the familiar distribution. Thus, generating observations from these conditional distributions is straightforward and fast. However, the conditional distributions $p(\xi_{i}|y, r, \theta, \phi)$ given in (8), $p(y_{m|i}|y_{a}, r, \xi_{i}, \theta, \phi)$ given in (9), $p(u|Y, F, \Lambda, \Psi)$ given in (10), for general case $p(\psi_{k}^{-1}Y, F, \Lambda)$ given in (11), $p(\lambda_{k}|Y, F, \Psi, u)$ given in (12), and $p(\phi|Y, F, r)$ given in (17) are non-standard and complex distributions. In these cases, it is rather difficult to directly generate observations from these conditional distributions. The well-known MH algorithm [29,30] is employed to generate observations from these conditional distributions with the help of proposal distributions from which it is easy to sample. Following the rationale given in [31], we consider the following proposal distributions $N(0, \sigma_{\xi}^{2}\Omega_{\xi})$, $N(0, \sigma_{\psi}^{2}\Omega_{\psi})$, $N(0, \sigma_{\lambda}^{2}\Omega_{\lambda})N(0, \sigma_{\phi}^{2}\Omega_{\phi})$, and $N(0, \sigma_{\phi}^{2}\Omega_{\phi})$ for sampling $\xi_{i}, y_{m|i}, u, \psi_{k}^{-1}, \lambda_{k}$ and $\phi$, respectively; where $\Omega_{\xi}^{-1} = \Sigma_{\xi} + \Lambda^T\Psi\Lambda$ with $\Psi = \text{diag}(d(y_{1};\mu_{1})/(2\psi_{1}), \ldots, d(y_{n};\mu_{p})/(2\psi_{p}))$ in which $d(y;\mu) = \delta^{2}d(y;\mu)/\mu^{2}$, and

$$
\Sigma_{\xi} = \sum_{i=1}^{n} \frac{\partial \log p(y_{i}|\mu_{1})}{\partial \xi_{i}}, \ldots, \sum_{i=1}^{n} \frac{\partial \log p(y_{i}|\mu_{p})}{\partial \xi_{i}}.
$$

The new MH algorithm for sampling observations $\xi_{i}, y_{m|i}, u, \psi_{k}^{-1}, \lambda_{k}$ and $\phi$ from their corresponding conditional distributions is implemented as follows. At the $(t+1)$th iteration with current observations $\xi_{i}^{(t)}, y_{m|i}^{(t)}, u^{(t)}, \psi_{k}^{-1}^{(t)}, \lambda_{k}^{(t)}$ and $\phi^{(t)}$, new candidates $\xi_{i}, y_{m|i}, u, \psi_{k}^{-1}, \lambda_{k}$ and $\phi$ are generated from the following distributions $N(\xi_{i}^{(t)}, \sigma_{\xi}^{2}\Omega_{\xi})$, $N(\psi_{k}^{-1}^{(t)}, \sigma_{\psi}^{2}\Omega_{\psi})$, $N(\lambda_{k}^{(t)}, \sigma_{\lambda}^{2}\Omega_{\lambda})$, $N(\psi_{k}^{-1}^{(t)}, \sigma_{\phi}^{2}\Omega_{\phi})$, and $N(\phi^{(t)}, \sigma_{\phi}^{2}\Omega_{\phi})$, respectively. They are accepted with the following probabilities

$$
\begin{align*}
&\min \left\{ 1, \frac{p(\xi_{i}|y_{m|i}, r, \theta, \phi)}{p(\xi_{i}^{(t)}|y_{m|i}, r, \theta, \phi)} \right\}, \\
&\min \left\{ 1, \frac{p(y_{m|i}|y_{a}, r, \xi_{i}, \theta, \phi)}{p(y_{m|i}^{(t)}|y_{a}, r, \xi_{i}, \theta, \phi)} \right\}, \\
&\min \left\{ 1, \frac{p(\lambda_{k}|Y, F, A, \Psi)}{p(\lambda_{k}^{(t)}|Y, F, A, \Psi)} \right\}, \\
&\min \left\{ 1, \frac{p(\phi|Y, F, r)}{p(\phi^{(t)}|Y, F, r)} \right\},
\end{align*}
$$

respectively.

3.3. Bayesian estimates and goodness-of-fit statistic

In this subsection, observations generated from the previous introduced algorithm are used to estimate latent variables $\xi_{i}$, unknown structural parameters in $\theta$ and $\phi$, and their standard errors. Also, we shall use these observations to construct a goodness-of-fit statistic.

Let $(F^{(t)}, \theta^{(t)}, \phi^{(t)}, Y_{m}^{(t)}): t = 1, \ldots, T$ be the random observations of $(F, \theta, \phi, Y_{m})$ generated from the joint conditional distribution $[F, \theta, \phi, Y_{m}|Y_{a}, r]$ via the above developed hybrid algorithm. The joint Bayesian estimates of $F, \theta, \phi$, and $Y_{m}$ are respectively given as:

$$
\hat{F} = \frac{1}{T} \sum_{t=1}^{T} F^{(t)}, \quad \hat{\theta} = \frac{1}{T} \sum_{t=1}^{T} \theta^{(t)}, \quad \hat{\phi} = \frac{1}{T} \sum_{t=1}^{T} \phi^{(t)}, \quad \hat{Y}_{m} = \frac{1}{T} \sum_{t=1}^{T} Y_{m}^{(t)}.
$$

These joint Bayesian estimates are consistent estimates of their corresponding posterior means [33]. Similarly, the sample covariance matrices of the generated observations can be used to estimate their corresponding posterior covariance matrices. For example, $\text{Var}(F|Y_{a}, r) = (T - 1)^{-1} \sum_{t=1}^{T} (F^{(t)} - \hat{F})(F^{(t)} - \hat{F})^{T}$ can be used as the estimate of $\text{Var}(F|Y_{a}, r)$. Thus, the diagonal elements of these matrices are just the estimates of the standard errors of their corresponding quantities.
To assess the plausibility of the posited model in Bayesian framework, Lee & Tang [18] extended the partial posterior predictive (PPP) p-value of Bayarri & Berger [20] to non-linear structural equation models. Similar to Lee & Tang [18], the PPP p-value for our considered models can be defined as

$$PPP_B = \int \text{pr}[D(Y_o|Y_m, F, \theta, \varphi, r) \geq D(Y_o^{\text{obs}}|Y_m, F, \theta, \varphi, r)]p^*(Y_m, F, \theta, \varphi)dY_mdFd\theta d\varphi.$$ 

(18)

in which $Y_o^{\text{obs}}$ is the observed value of manifest variable $Y_o$,

$$p^*(Y_m, F, \theta, \varphi) \propto \frac{p(Y_o^{\text{obs}}, Y_m, F, \theta, \varphi)p(\theta, \varphi)}{p(D^{\text{obs}}|Y_m, F, \theta, \varphi)}.$$ 

$D(\cdot)$ is a discrepancy variable, and $D^{\text{obs}}$ is the observed value of $D(\cdot)$ based on $Y_o^{\text{obs}}$. For our problem, we may choose the following discrepancy variable:

$$D(Y_o|Y_m, F, \theta, \varphi, r) = \sum_{i=1}^n (y_{oi} - u_{o}^* Y_i^{-1}(y_{oi} - u_{oi}^*),$$

where $u_{o}^*$ and $V_o$ are conditional expectation and conditional covariance of $y_{oi}$ given $Y_m, F, \theta, \varphi$ and $r$. From the definition of $D(\cdot)$, it is easily shown that the distribution of this discrepancy variable is asymptotically distributed as chi-square with degree of freedom $\sum_{i=1}^n P_{ii}$ which indicates that the distribution of $p(D^{\text{obs}}|Y_m, F, \theta, \varphi)$ is independent of $(Y_m, F, \theta, \varphi)$. Thus, we have

$$p^*(Y_m, F, \theta, \varphi) \propto \frac{p(Y_o^{\text{obs}}, Y_m, F, \theta, \varphi)p(\theta, \varphi)}{}.$$ 

It is easily seen from (18) that it is rather difficult to obtain the PPP p-value because of high dimensional integral involved. Hence, the Monte Carlo method is used to obtain the above difficulties. Let $\{(Y_m^{(t)}, F^{(t)}, \theta^{(t)}, \varphi^{(t)}): t = 1, \ldots, T\}$ be the observations generated from $p^*(Y_m, F, \theta, \varphi)$, and $\{Y_o^{(t)}: t = 1, \ldots, T\}$ be the observations generated from $p(Y_o|Y_m, F, \theta)$. Similar to [18], the PPP p-value can be estimated by

$$\hat{PPP}_B = T^{-1} \sum_{j=1}^T I(D(Y_o^{(j)}|Y_m^{(j)}, F^{(j)}, \theta^{(j)}, \varphi^{(j)}, r) \geq D(Y_o^{\text{obs}}|Y_m^{(j)}, F^{(j)}, \theta^{(j)}, \varphi^{(j)}, r)).$$

where $I(\cdot)$ is an indicator function. The above $\hat{PPP}_B$-value can be calculated via the following steps:

Step 1. Sample observations $(Y_m^{(t)}, F^{(t)}, \theta^{(t)}, \varphi^{(t)})$ from $p^*(Y_m, F, \theta, \varphi)$ by using the algorithm developed in Sections 3.1 and 3.2.

Step 2. Generate $Y_o^{(t)}$ from $p(Y_o|Y_m^{(t)}, F^{(t)}, \theta^{(t)}) = \prod_{i=1}^n p(y_{oi}|Y_m^{(t)}, \xi^{(t)}, \theta^{(t)})$ and compute discrepancies $D(Y_o^{(t)}|Y_m^{(t)}, F^{(t)}, \theta^{(t)}, \varphi^{(t)}, r)$ and $D(Y_o^{\text{obs}}|Y_m^{(t)}, F^{(t)}, \theta^{(t)}, \varphi^{(t)}, r)$. If the former is larger than or equal to the latter, let $\omega^{(t)} = 1$; otherwise, $\omega^{(t)} = 0$.

Step 3. Update $t$, repeat Step 1 and Step 2. Then, $\hat{PPP}_B = T^{-1} \sum_{j=1}^T \omega^{(j)}$.

Although the specified missingness mechanism itself is not “testable” [34], we could compare any two different models via Bayes factor. Now we consider the extension of the well-known Bayes factor for model comparisons to our current considered models. Let $M_0$ and $M_1$ be two competing models, the Bayes factor [35] is defined as:

$$B_{10} = \frac{p(Y_o, r|M_1)}{p(Y_o, r|M_0)},$$

where

$$p(Y_o, r|M_k) = \int p(Y_o, Y_m, F, r|\theta_k, \varphi_k, M_k)p(\theta_k, \varphi_k|M_k)dY_mdFd\theta_k d\varphi_k, \quad k = 1, 0,$$

is the marginal density of $M_k$ with parameter vectors $\theta_k$ and $\varphi_k$. Clearly, it is rather difficult to obtain $p(Y_o, r|M_k)$ because of high dimensional integral involved. Here, a procedure is presented to calculate logarithm Bayes factor on the basis of path sampling of Gelman & Meng [21].

Following [14], we consider the following class of densities

$$z(t) = p(Y_o, r|t) = \int p(Y_m, Y_o, F, r|\theta, \varphi, t)dY_mdFd\theta d\varphi = \int p(Y_m, Y_o, F, r|\theta, \varphi)p(\theta, \varphi)d\theta d\varphi,$$

where $t$ is a continuous parameter belonging to interval $[0, 1]$, and $p(Y_m, Y_o, F, r|\theta, \varphi)$ is the density of model $M_t$ that links $M_0$ and $M_1$ with the continuous parameter $t$, such that $M_t = M_0$ if $t = 0$ and $M_t = M_1$ if $t = 1$. Following [21], we have

$$\log B_{10} = \log \frac{z(1)}{z(0)} = \int_0^1 E_{Y_m, F, \theta, \varphi} U(Y_m, Y_o, F, r|\theta, \varphi, t)dt,$$
where \( U(Y_m, Y_o, F, r, \theta, \varphi, t) = \frac{d \log p(Y_m, Y_o, F, r, t|\theta, \varphi)}{dt} \), and \( E_{Y_m,F,\theta,\varphi} \) is the expectation with respect to the distribution \( p(Y_m, F, \theta, \varphi|Y_o, r, t) \). Let \( 0 = t_0 < t_1 < \cdots < t_{(3)} < t_{(5)} = 1 \). Then, \( \log B_{10} \) can be estimated by

\[
\log B_{10} = \frac{1}{2} \sum_{j=0}^{5} (t_{(j+1)} - t_{(j)}) (\hat{U}_{(j+1)} + \hat{U}_{(j)}),
\]

where

\[
\hat{U}_{(j)} = J^{-1} \sum_{j=1}^{I} U(Y^{(i)}_m, Y_o, F^{(i)}, r, \theta^{(i)}, \varphi^{(i)}, t_{(j)}),
\]

and \( \{ (Y^{(i)}_m, F^{(i)}, \theta^{(i)}, \varphi^{(i)}) ; j = 1, \ldots, J \} \) are observations generated from \( p(Y_m, F, \theta, \varphi|Y_o, r, t_{(j)}) \propto \prod_{i=1}^{I} p(y_i|\xi_i, \theta, t_{(j)}) p(\xi_i|\theta, t_{(j)}) p(r|y_i, \xi_i, \theta, t_{(j)}) p(\theta, \varphi) \) which can be implemented via the hybrid algorithm introduced in Sections 3.1 and 3.2. In the simulation study, we take \( S = 9 \) and \( J = 2000 \) after a burn-in of 3000 iterations.

### 4. Numerical examples

In this section, a simulation study and a real example are used to show the above proposed Bayesian approach.

#### 4.1. Simulation studies

In this subsection, a simulation study is used to investigate the sensitivity of the Bayesian estimates with respect to prior inputs and the choice of the missingness mechanism. To address the above issues, a data set \( \{ y_i : i = 1, \ldots, n \} \) is generated from a non-linear structural equation model with nine manifest variables from a reproductive dispersion model in which the nine manifest variables are related with three basic factors \( \xi_{(1)} = \xi_{(1)}, \xi_{(2)} = (\xi_{(2)}, \xi_{(3)})^T \). For a measurement model defined in (3), \( y_i \) are generated from the following Gumbel distribution with parameter \( \mu_i \), i.e.,

\[
p(y_i|\xi_i) = \exp(y_i - \mu_i - \exp(y_i - \mu_i)), \quad i = 1, \ldots, n; \ j = 1, \ldots, 9;
\]

where \( \mu_i = u_i + \lambda_i \xi_i \). In this simulation study, we set the true values of the unknown parameters to be \( u_1 = \cdots = u_9 = 0 \), \( \lambda_1 = \lambda_3 = \lambda_5 = -0.5, \lambda_2 = \lambda_4 = 0.6, \lambda_6 = \lambda_7 = 1.0, \gamma_1 = \gamma_2 = 0.56, \gamma_3 = 0.36, \gamma_4 = 1.0 \).

Missing data are generated from the following missingness mechanism model

\[
M_0 : m_0(y_i, \xi_i, \varphi) = \log(\text{pr}(r_{ij} = 1|y_i, \xi_i, \varphi)) = \varphi_0 + \varphi_1 y_{ij} + \cdots + \varphi_9 y_{ij}
\]

with true parameter \( \varphi_0 = -2.7, \varphi_1 = \cdots = \varphi_9 = -0.3 \). The missing data are created as follows: (a) generate the data set \( \{ y_i : i = 1, \ldots, n; j = 1, \ldots, 9 \} \) from the above Gumbel distribution given in (19); (b) determine whether the observation \( y_{ij} \) is missing or not via the missing mechanism model given in (20) with the above true values of \( \varphi_0, \varphi_1, \ldots, \varphi_9 \).

More specifically, we generate a random number \( \kappa \) from the uniform distribution \( U(0, 1) \), the observation \( y_{ij} \) is missing if \( \kappa \leq \text{pr}(r_{ij} = 1|y_i, \xi_i, \varphi) = 1/(1 + \exp(-\varphi_0 - \varphi_1 y_{ij} - \cdots - \varphi_9 y_{ij})) \). There are 32 unknown parameters in the above specified model. The average proportion of missing data generated in this way is about 0.43. Bayesian estimates of the unknown parameters are obtained on the basis of 100 replications for \( n = 300 \).

To investigate the sensitivity of the Bayesian estimates to prior inputs, we consider the following two kinds of hyper-parameters. Type I: the hyper-parameters for \( \alpha_0, \lambda_{01}, \lambda_{02} = I_0 \) and \( \varphi_0 \) are taken to be their corresponding true values, \( \Sigma_0 = 0.25I, \text{Var}_k = 0.25, \text{Var}_{k1} = 0.25I, \text{Var}_0 = 0.5 \Phi_0, \alpha_{01} = 10, \beta_{01} = 8, V = 1 \). This can be regarded as a situation with good prior information. Type II: non-informative prior. Results which are obtained under different types of prior inputs via the non-ignorable missingness mechanism model (7) are reported in Table 1, where ‘Bias’ denotes the absolute difference between the true value and the mean of the estimates, and ‘RMS’ is the root mean square between the estimates and their true values. From Table 1, we observe that the Bayesian estimates are reasonably accurate under different prior inputs, and are not sensitive to prior inputs.

Now we investigate the influence of the missingness mechanism model on the Bayesian estimates. In this study, complete data are generated on the basis of the above specified model and true parameters. But, missing data are created via the following three types of missingness mechanisms.

Type A. non-ignorable missingness mechanism which is different from \( M_0 \) given in (20):

\[
\log(\text{pr}(r_{ij} = 1|y_i, \xi_i, \varphi)) = \varphi_0 + \varphi_1 y_{ij} + \cdots + \varphi_9 y_{ij} + \varphi_{10} \xi_{1j} + \varphi_{11} \xi_{2j} + \varphi_{12} \xi_{3j}
\]

with \( \varphi_0 = -2.0, \varphi_1 = \cdots = \varphi_{12} = 0.1 \).
Table 2
Performance of the Bayesian estimates in the simulation study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type I</th>
<th>Type II</th>
<th>Par.</th>
<th>Type I</th>
<th>Type II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMS</td>
<td>Bias</td>
<td>RMS</td>
<td>Bias</td>
</tr>
<tr>
<td>$u_1$</td>
<td>0.007</td>
<td>0.056</td>
<td>0.015</td>
<td>0.075</td>
<td>0.068</td>
</tr>
<tr>
<td>$u_2$</td>
<td>0.006</td>
<td>0.061</td>
<td>0.004</td>
<td>0.070</td>
<td>0.026</td>
</tr>
<tr>
<td>$u_3$</td>
<td>0.008</td>
<td>0.117</td>
<td>0.018</td>
<td>0.113</td>
<td>0.010</td>
</tr>
<tr>
<td>$u_4$</td>
<td>0.006</td>
<td>0.083</td>
<td>0.007</td>
<td>0.094</td>
<td>0.031</td>
</tr>
<tr>
<td>$u_5$</td>
<td>0.004</td>
<td>0.074</td>
<td>0.013</td>
<td>0.081</td>
<td>0.031</td>
</tr>
<tr>
<td>$u_6$</td>
<td>0.011</td>
<td>0.093</td>
<td>0.018</td>
<td>0.095</td>
<td>0.022</td>
</tr>
<tr>
<td>$u_7$</td>
<td>0.017</td>
<td>0.114</td>
<td>0.014</td>
<td>0.098</td>
<td>0.049</td>
</tr>
<tr>
<td>$u_8$</td>
<td>0.005</td>
<td>0.068</td>
<td>0.009</td>
<td>0.071</td>
<td>0.036</td>
</tr>
<tr>
<td>$u_9$</td>
<td>0.013</td>
<td>0.132</td>
<td>0.035</td>
<td>0.134</td>
<td>0.023</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.021</td>
<td>0.095</td>
<td>0.026</td>
<td>0.104</td>
<td>0.026</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.026</td>
<td>0.087</td>
<td>0.020</td>
<td>0.090</td>
<td>0.054</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.035</td>
<td>0.072</td>
<td>0.019</td>
<td>0.085</td>
<td>0.018</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.043</td>
<td>0.116</td>
<td>0.056</td>
<td>0.130</td>
<td>0.026</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.014</td>
<td>0.070</td>
<td>0.015</td>
<td>0.085</td>
<td>0.079</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>0.046</td>
<td>0.108</td>
<td>0.036</td>
<td>0.134</td>
<td>0.079</td>
</tr>
<tr>
<td>$\gamma_5$</td>
<td>0.023</td>
<td>0.082</td>
<td>0.017</td>
<td>0.190</td>
<td>0.094</td>
</tr>
<tr>
<td>$\gamma_6$</td>
<td>0.057</td>
<td>0.071</td>
<td>0.049</td>
<td>0.088</td>
<td>0.079</td>
</tr>
<tr>
<td>$\gamma_7$</td>
<td>0.039</td>
<td>0.087</td>
<td>0.033</td>
<td>0.093</td>
<td>0.085</td>
</tr>
</tbody>
</table>

Type B. MAR missingness mechanism.

Type C. Logistic regression model given in above $M_0$.

All Bayesian estimates of unknown parameters are obtained via type I prior inputs. For missing data generated via the first two missingness mechanism models, Bayesian estimates are calculated on the basis of the incorrect missingness mechanism model $M_0$. For missing data generated via the incorrect MAR missingness mechanism, these Bayesian estimates are presented in Table 2. Examination of Table 2 reveals that (1) Bayesian estimates obtained using the model $M_0$ are quite accurate when the true missingness mechanisms are more complicated models or the true missing data are MAR; (2) Bayesian estimates obtained under the incorrect MAR assumption are inaccurate. These findings are consistent with those given in [19].

To illustrate application of the path sampling procedure in comparing various missingness mechanism models, we consider the following three logistic models:

$M_0 : m_0(y_i, \xi_i, \varphi) = \varphi_0 + \varphi_1 y_{i1} + \cdots + \varphi_9 y_{i9},$

$M_1 : m_1(y_i, \xi_i, \varphi) = \varphi_0 + \varphi_{10} \xi_{i1} + \varphi_{11} \xi_{i2} + \varphi_{12} \xi_{i3},$

$M_2 : m_2(y_i, \xi_i, \varphi) = \varphi_0 + \varphi_{10} y_{i1} + \cdots + \varphi_9 y_{i9} + \varphi_{10} \xi_{i1} + \varphi_{11} \xi_{i2} + \varphi_{12} \xi_{i3}.$

In this study, the complete data set is generated with the above specified model (19) and the corresponding true parameters, and the missing data are generated via model $M_0$. Similar to [19], the linking model of $M_0$ and $M_1$ is given by

$M_{01} : m_{01}(y_i, \xi_i, \varphi) = \varphi_0 + (1 - t) \varphi_1 y_{i1} + \cdots + (1 - t) \varphi_9 y_{i9} + t \varphi_{10} \xi_{i1} + t \varphi_{11} \xi_{i2} + t \varphi_{12} \xi_{i3},$

and the linking model of $M_0$ and $M_2$ is given by

$M_{02} : m_{02}(y_i, \xi_i, \varphi) = \varphi_0 + \varphi_1 y_{i1} + \cdots + \varphi_9 y_{i9} + t \varphi_{10} \xi_{i1} + t \varphi_{11} \xi_{i2} + t \varphi_{12} \xi_{i3}.$

Clearly, $M_{01}$ is equal to $M_0$ or $M_1$ when $t = 0$ or 1, and $M_{02}$ is equal to $M_0$ or $M_2$ when $t = 0$ or 1. The log Bayes factor obtained by using the path sampling procedure and hyper-parameters given in Type I and Type II are log $B_{01} = -14.26$ and $-18.733$, and log $B_{02} = -21.375$ and $-37.893$, respectively. Based on the criterion of model comparison of [32], the true model $M_0$ is selected, which implies that the proposed model comparison procedure is rather effective.

4.2. Real examples

In this example, a small portion of the ICPSR data set collected by the World Values Survey 1981–1984 and 1990–1993 [38] is used to illustrate the proposed methodologies. The whole data set was collected in 50 societies around the world on broad topics such as work, religious belief, the meaning and purpose of life, family life, contemporary social issues, etc. Here, only data from the females in Russia are used, and the ICPSR data set has been analyzed by Lee & Tang [18]. Variables 116, 117, 252, 253, 254, 296, 298 and 314 in original data set are taken to be manifest variables which form $y = (y_1, \ldots, y_8)$. These variables are measured on a 10-point scale, for brevity, we regarded them as continuous. There are 1124 random observations and 111 different missing patterns in the data set in which there are only 451 (40.12%) fully observed cases. The manifest variable $y_1$ is missing most often, and its missing frequency is 255. It is easily seen from Table 3 and the meanings of variables that the unanswered questions are either related to personal attitudes or related to personal morality. Thus, it is reasonable to assume that their corresponding missing data are non-ignorable.
Based on the meanings of the questions corresponding to the manifest variables, we assume that $y_{ij}$ are distributed as the following conditional Gumbel distribution with parameter $\mu_{ij}$, i.e.,

$$ p(y_{ij}) = \exp\{y_{ij} - \mu_{ij} - \exp(y_{ij} - \mu_{ij})\}, \quad i = 1, \ldots, 1124, j = 1, \ldots, 8, $$

and $\mu_i = (\mu_{i1}, \ldots, \mu_{i8})^T = u + A \xi_i$ where $u = (u_1, \ldots, u_8)^T$, $\xi_i = (\xi_{i1}, \xi_{i2}, \xi_{i3})^T$ in which $\xi_{i1}$ corresponds to the first two manifest variables $(y_{i1}, y_{i2})$ which can be roughly interpreted as “job satisfaction”, $\xi_{i2}$ corresponds to the next three manifest variables $(y_{i3}, y_{i4}, y_{i5})$ which can be interpreted as “job attitude”, $\xi_{i3}$ corresponds to the last three manifest variables $(y_{i6}, y_{i7}, y_{i8})$ which can be interpreted as “morality (in relation to money)”; and

$$ A^T = \begin{pmatrix} 1.0^* & \lambda_{21} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 1.0^* & \lambda_{42} & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{73} & \lambda_{83} & 0.0^* \end{pmatrix}. $$

Also, we consider the following non-linear structural equation model:

$$ M_1 : \xi_{i1} = \gamma_{11} \xi_{i2} + \gamma_{21} \xi_{i3} + \gamma_{31} \xi_{i4} + \delta_i, $$

where $\delta_i \sim N(0, \psi_3)$. $\xi_{i(2)} = (\xi_{i2}, \xi_{i3})^T \sim N(0, \Phi)$ with

$$ \Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}. $$

The path sampling procedure is used to calculate the Bayes factor for comparing the model $M_M$: MAR missingness mechanism with the following two non-ignorable missingness mechanism models:

$$ M_A : \logit\{pr(r_{ij} = 1|y_{ij}, \xi_i, \varphi)\} = \varphi_0 + \varphi_1 y_{i1} + \cdots + \varphi_8 y_{i8}; $$

$$ M_B : \logit\{pr(r_{ij} = 1|y_{ij}, \xi_i, \varphi)\} = \varphi_0 + \varphi_1 \xi_{i1} + \varphi_2 \xi_{i2} + \varphi_3 \xi_{i3}. $$

Note that $M_A$ only considers all the manifest variables, whilst $M_B$ only considers all the latent variables. The Bayes factors for comparing the above models are computed via the following selected prior inputs: $\alpha_{031} = 10, \beta_{031} = 4, \theta_0 = 5 \Phi$, $H_0 = 2I, H_k = 2I, u_k = \tilde{u}, \lambda_{0k} = \tilde{\lambda}_k, \lambda_{0k1} = \tilde{\Gamma}_k, \varphi^0 = \tilde{\varphi}, V = 2I, \Sigma_0 = I$, where $\tilde{\Phi}, \tilde{\varphi}, \tilde{\lambda}_k, \tilde{\Gamma}_k, \tilde{\varphi}$ are the Bayesian estimates obtained by analyzing the model $M_1$ with non-informative prior assumptions. The number of grids in the path

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type A Bias</th>
<th>RMS</th>
<th>Type A Bias</th>
<th>RMS</th>
<th>Type A Bias</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1$</td>
<td>0.015</td>
<td>0.130</td>
<td>0.013</td>
<td>0.120</td>
<td>0.232</td>
<td>0.259</td>
</tr>
<tr>
<td>$u_2$</td>
<td>0.011</td>
<td>0.113</td>
<td>0.019</td>
<td>0.108</td>
<td>0.150</td>
<td>0.207</td>
</tr>
<tr>
<td>$u_3$</td>
<td>0.013</td>
<td>0.127</td>
<td>0.016</td>
<td>0.147</td>
<td>0.256</td>
<td>0.285</td>
</tr>
<tr>
<td>$u_4$</td>
<td>0.033</td>
<td>0.088</td>
<td>0.028</td>
<td>0.089</td>
<td>0.382</td>
<td>0.389</td>
</tr>
<tr>
<td>$u_5$</td>
<td>0.004</td>
<td>0.085</td>
<td>0.011</td>
<td>0.093</td>
<td>0.413</td>
<td>0.440</td>
</tr>
<tr>
<td>$u_6$</td>
<td>0.013</td>
<td>0.082</td>
<td>0.034</td>
<td>0.083</td>
<td>0.223</td>
<td>0.274</td>
</tr>
<tr>
<td>$u_7$</td>
<td>0.020</td>
<td>0.073</td>
<td>0.022</td>
<td>0.070</td>
<td>0.368</td>
<td>0.373</td>
</tr>
<tr>
<td>$u_8$</td>
<td>0.021</td>
<td>0.098</td>
<td>0.045</td>
<td>0.100</td>
<td>0.351</td>
<td>0.376</td>
</tr>
<tr>
<td>$u_9$</td>
<td>0.035</td>
<td>0.131</td>
<td>0.033</td>
<td>0.156</td>
<td>0.316</td>
<td>0.335</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.016</td>
<td>0.043</td>
<td>0.009</td>
<td>0.042</td>
<td>0.053</td>
<td>0.093</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.013</td>
<td>0.052</td>
<td>0.027</td>
<td>0.056</td>
<td>0.036</td>
<td>0.116</td>
</tr>
<tr>
<td>$\lambda_{21}$</td>
<td>0.027</td>
<td>0.037</td>
<td>0.049</td>
<td>0.072</td>
<td>0.121</td>
<td>0.175</td>
</tr>
<tr>
<td>$\lambda_{22}$</td>
<td>0.019</td>
<td>0.067</td>
<td>0.037</td>
<td>0.083</td>
<td>0.085</td>
<td>0.154</td>
</tr>
<tr>
<td>$\lambda_{31}$</td>
<td>0.008</td>
<td>0.088</td>
<td>0.009</td>
<td>0.052</td>
<td>0.049</td>
<td>0.121</td>
</tr>
<tr>
<td>$\lambda_{93}$</td>
<td>0.014</td>
<td>0.112</td>
<td>0.011</td>
<td>0.143</td>
<td>0.117</td>
<td>0.198</td>
</tr>
<tr>
<td>$\psi_0$</td>
<td>0.027</td>
<td>0.148</td>
<td>0.054</td>
<td>0.168</td>
<td>0.335</td>
<td>0.335</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>0.018</td>
<td>0.113</td>
<td>0.019</td>
<td>0.118</td>
<td>0.118</td>
<td>0.118</td>
</tr>
<tr>
<td>$\psi_2$</td>
<td>0.011</td>
<td>0.161</td>
<td>0.032</td>
<td>0.157</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>$\psi_3$</td>
<td>0.015</td>
<td>0.102</td>
<td>0.016</td>
<td>0.116</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>$\psi_4$</td>
<td>0.040</td>
<td>0.116</td>
<td>0.043</td>
<td>0.129</td>
<td>0.043</td>
<td>0.129</td>
</tr>
<tr>
<td>$\psi_5$</td>
<td>0.036</td>
<td>0.112</td>
<td>0.033</td>
<td>0.112</td>
<td>0.033</td>
<td>0.112</td>
</tr>
<tr>
<td>$\psi_6$</td>
<td>0.021</td>
<td>0.127</td>
<td>0.028</td>
<td>0.125</td>
<td>0.028</td>
<td>0.125</td>
</tr>
<tr>
<td>$\psi_7$</td>
<td>0.015</td>
<td>0.126</td>
<td>0.021</td>
<td>0.127</td>
<td>0.021</td>
<td>0.127</td>
</tr>
<tr>
<td>$\psi_8$</td>
<td>0.012</td>
<td>0.114</td>
<td>0.013</td>
<td>0.117</td>
<td>0.013</td>
<td>0.117</td>
</tr>
<tr>
<td>$\psi_9$</td>
<td>0.019</td>
<td>0.145</td>
<td>0.024</td>
<td>0.172</td>
<td>0.024</td>
<td>0.172</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.025</td>
<td>0.146</td>
<td>0.019</td>
<td>0.149</td>
<td>0.374</td>
<td>0.385</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.016</td>
<td>0.123</td>
<td>0.032</td>
<td>0.128</td>
<td>0.108</td>
<td>0.124</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.034</td>
<td>0.222</td>
<td>0.067</td>
<td>0.250</td>
<td>0.067</td>
<td>0.206</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>0.032</td>
<td>0.097</td>
<td>0.051</td>
<td>0.104</td>
<td>0.135</td>
<td>0.170</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>0.037</td>
<td>0.072</td>
<td>0.041</td>
<td>0.169</td>
<td>0.301</td>
<td>0.352</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>0.017</td>
<td>0.103</td>
<td>0.114</td>
<td>0.111</td>
<td>0.436</td>
<td>0.442</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>0.040</td>
<td>0.127</td>
<td>0.054</td>
<td>0.136</td>
<td>0.224</td>
<td>0.232</td>
</tr>
</tbody>
</table>
Table 3
Missing patterns and their sample sizes: ICPSR data, “x” and “o” indicate missing and observed entries, respectively.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Manifest variables</th>
<th>Sample size</th>
<th>Pattern</th>
<th>Manifest variables</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>o o o o o o o</td>
<td>451</td>
<td>2</td>
<td>x o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>x o o o o o o</td>
<td>40</td>
<td>3</td>
<td>o x o o o o o</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>o x o o o o o</td>
<td>41</td>
<td>4</td>
<td>o o o o o o o</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>o o o o o o o</td>
<td>42</td>
<td>5</td>
<td>o x o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>o o o o o o o</td>
<td>43</td>
<td>6</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>o o o o o o o</td>
<td>44</td>
<td>7</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>o o o o o o o</td>
<td>45</td>
<td>8</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>o o o o o o o</td>
<td>46</td>
<td>9</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>o o o o o o o</td>
<td>47</td>
<td>10</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>o o o o o o o</td>
<td>48</td>
<td>12</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>o o o o o o o</td>
<td>49</td>
<td>14</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>o o o o o o o</td>
<td>50</td>
<td>16</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>o o o o o o o</td>
<td>51</td>
<td>18</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>o o o o o o o</td>
<td>52</td>
<td>20</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>o o o o o o o</td>
<td>53</td>
<td>22</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>o o o o o o o</td>
<td>54</td>
<td>24</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>o o o o o o o</td>
<td>55</td>
<td>26</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>o o o o o o o</td>
<td>56</td>
<td>28</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>o o o o o o o</td>
<td>57</td>
<td>30</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>o o o o o o o</td>
<td>58</td>
<td>32</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>o o o o o o o</td>
<td>59</td>
<td>34</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>o o o o o o o</td>
<td>60</td>
<td>36</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td>o o o o o o o</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Manifest variables</th>
<th>Sample size</th>
<th>Pattern</th>
<th>Manifest variables</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>o o o o o o o</td>
<td>62</td>
<td>2</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>o o o o o o o</td>
<td>63</td>
<td>4</td>
<td>o o o o o o o</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>o o o o o o o</td>
<td>64</td>
<td>6</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>o o o o o o o</td>
<td>65</td>
<td>8</td>
<td>o o o o o o o</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>o o o o o o o</td>
<td>66</td>
<td>10</td>
<td>o o o o o o o</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>o o o o o o o</td>
<td>67</td>
<td>12</td>
<td>o o o o o o o</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>o o o o o o o</td>
<td>68</td>
<td>14</td>
<td>o o o o o o o</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>o o o o o o o</td>
<td>69</td>
<td>16</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>o o o o o o o</td>
<td>70</td>
<td>18</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>o o o o o o o</td>
<td>71</td>
<td>20</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>o o o o o o o</td>
<td>72</td>
<td>22</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>o o o o o o o</td>
<td>73</td>
<td>24</td>
<td>o o o o o o o</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>o o o o o o o</td>
<td>74</td>
<td>26</td>
<td>o o o o o o o</td>
<td>2</td>
</tr>
</tbody>
</table>

The estimated log Bayes factors are equal to $\log B_{BA} = -89.75$, and $\log B_{P=-56.26}$, which indicate that the non-ignorable missingness mechanism defined in $M_A$ is better than that defined in $M_B$, and is also better than the ignorable missingness mechanism defined in $M_{BA}$. The PPP $p$-value corresponding to the non-linear structural equation model $M_1$ with the non-ignorable missingness mechanism defined in model $M_A$ is equal to 0.398, which indicates that the proposed non-linear structural equation model and the selected non-ignorable missingness mechanism model are plausible.

The Bayesian estimates and their standard error estimates of the unknown parameters in the selected model are presented in Table 4. Examination of Table 4 reveals that: (1) the estimates of the coefficients $\phi_0$, $\phi_2$, $\phi_3$ and $\phi_4$ are significantly different from zero, which further indicates that our considered non-ignorable missingness mechanism defined in model $M_A$ is necessary. This result is consistent with the conclusion that was obtained by model comparison. (2) the estimates of the loading factors are rather large, which indicates a strong associations between the latent variables and their corresponding indicators. To save space, the less interesting estimates obtained from $M_B$ are not presented. By values of $\hat{\phi}_{(1)}$, $\hat{\phi}_{(2)}$, and $\hat{\phi}_{(22)}$, we observe that the estimate of the correlation coefficient between $\xi_2$ and $\xi_3$ is 0.07, which indicates that “job attitude” $\xi_2$ and “morality” $\xi_3$ is weakly correlated. The estimated non-linear structural equation $\xi_1 = 0.098\xi_2 + 0.640\xi_3 - 0.763\xi_2 \xi_3$ has the following interpretations: (i) $\hat{\gamma}_1 = -0.098$ indicates that better job satisfaction (negative $\xi_2$) has a positive linear impact on job satisfaction; (ii) $\hat{\gamma}_2 = 0.640$ implies that this lower moral standard (positive $\xi_3$) has a positive linear impact on job satisfaction; (iii) $\hat{\gamma}_3 = -0.763$ reveals that $\xi_2$ and $\xi_3$ have an interaction effect on job satisfaction.

5. Discussion

Reproductive dispersion model [10,8] includes a wide variety of distributions such as Normal, Exponential family, Binomial, Gamma, Gumbel and simplex distributions as its special case. Based on the reproductive dispersion model,
some new models are proposed, for example, proper dispersion models [36], marginal models for longitudinal continuous proportional data [13], non-linear reproductive dispersion models [11,12], non-linear reproductive dispersion mixed models [37]. However, there is little work done in non-linear structural equation models with manifest variables from a reproductive dispersion model. Hence, we extended work of Lee & Tang [4] for non-linear structural equation models with manifest variables from an exponential family to non-linear structural equation models with manifest variables from a reproductive dispersion model, and we also consider the case that manifest variables have non-ignorable missing data. This is one of main contribution of our paper.

In this paper, a Bayesian procedure for analyzing non-linear structural equation model with manifest variables from a reproductive dispersion model and having non-ignorable missing data is developed. On the basis of a non-ignorable missing data model that is formulated via a logistic regression model, recent developed tools in statistics computing, such as the Gibbs sampler, the M–H algorithm and path sampling are used to calculate the Bayesian estimates of structural parameter and their corresponding standard error estimates, the PPP p-value for goodness-of-fit and the Bayesian factor for model comparison. From the simulation results given in Section 4.1, we observe that the non-ignorable missing data can not be treated as MAR (Missing at random), and the newly developed statistical methods for analyzing non-ignorable missing data are necessary. Moreover, application of above statistical computing tools to the proposed non-linear structural equation model with manifest variables from a reproductive dispersion model and/or having non-ignorable missing data are novel and non–trivial.

References