数理解析研究所講究録
第1703巻 2010年 112-133

Analysis of the Latent Growth Model with Dropout Data: A Bayesian Perspective

Daisuke TANAKA and Yuichiro KANAZAWA

Graduate School of Systems and Information Engineering
University of Tsukuba

1 Model

1.1 Latent Growth Model

Introduction

The LGM examines the development of individuals on one or more outcome variables over time. In growth modeling, random effects are used to capture individual differences in development. The LGM comes from the structural equation model (SEM) approach in that random effects to determine the trajectory are treated as latent variables. Covariates affecting these latent variables are included in this model. By doing so, we understand which covariates affect the development of subjects' outcomes through random effects. This means that individuals with differing covariate values exhibit differing growth curves. This model gives us a tool to examine complicated relationships between covariates affecting latent growth factors. Figure 1 shows path diagram of the LGM with time invariant covariates.

Model Specification

Suppose $\boldsymbol{y}_i$ is the outcome for subject i, and it satisfies the following measurement equation:

$$\boldsymbol{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \tag{1.1}$$

where $\boldsymbol{y}_i$ is a $T \times 1$ vector of observable outcomes (that is, both observed and missing included), $\boldsymbol{\Lambda}$ is a $T \times M$ matrix of coefficients with the first column often defined to be unity, $\boldsymbol{\eta}_i$ is a $M \times 1$ vector of latent growth factors for subject i, and $\boldsymbol{\epsilon}_i$ is a $T \times 1$ vector of measurement errors associated with $\boldsymbol{y}_i$.

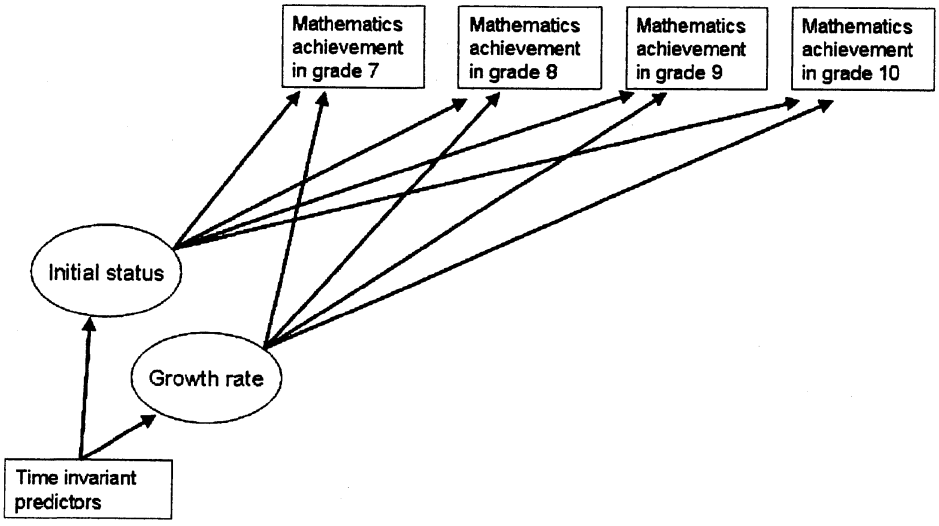

Figure 1: Path diagram of the LGM. The manifest variables are in box and the latent variables are in oval.

The elements of (1.1) are

$$\begin{pmatrix} y_{1i} \\ \vdots \\ y_{ti} \\ \vdots \\ y_{Ti} \end{pmatrix} = \begin{pmatrix} \lambda_{10} & \cdots & \lambda_{1m} & \cdots & \lambda_{1,M-1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \lambda_{t0} & \cdots & \lambda_{tm} & \cdots & \lambda_{t,M-1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \lambda_{T0} & \cdots & \lambda_{Tm} & \cdots & \lambda_{T,M-1} \end{pmatrix} \begin{pmatrix} \eta_{0i} \\ \vdots \\ \eta_{mi} \\ \vdots \\ \eta_{M-1,i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \vdots \\ \epsilon_{ti} \\ \vdots \\ \epsilon_{Ti} \end{pmatrix}. \tag{1.2}$$

The $\mathbf{\Lambda}$ denotes the time score which reflects the numerical value of "time," but is measured in units depending upon situations (e.g., seconds, minutes, decades, or grades in middle to high schools).

For a simple linear trajectory model, the time score is measured equidistantly and as such is defined to be $\lambda_{t0} = 1$ and $\lambda_{t1} = t-1$, where $t = 1, \ldots, T$, otherwise is 0, where T represents the number of time points in the elements notation (1.2). Intercept growth factor η_{0i} measures systematic part of the variation in the outcome variable at the time point where the time score is zero. Slope growth factor η_{1i} measures systematic part of the increase in the outcome variable for a time score increase of one unit. The elements of the

simple linear trajectory model are presented as

$$\begin{pmatrix} y_{1i} \\ \vdots \\ y_{ti} \\ \vdots \\ y_{Ti} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & t \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix} \begin{pmatrix} \eta_{0i} \\ \eta_{1i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \vdots \\ \epsilon_{ti} \\ \vdots \\ \epsilon_{Ti} \end{pmatrix}.$$

Suppose the latent growth factors $(\eta_{0i}, \cdots, \eta_{M-1,i})^T$ satisfy the following structural equation:

$$\boldsymbol{\eta}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{x}_i + \boldsymbol{\zeta}_i, \tag{1.3}$$

where $\boldsymbol{\mu}$ is a $M \times 1$ vector of intercepts across all subjects, $\boldsymbol{\Gamma}$ is a $M \times K$ matrix of coefficients, $\boldsymbol{x}_i$ is a $K \times 1$ vector of observed time invariant covariates, and $\boldsymbol{\zeta}_i$ is a $M \times 1$ vector of errors or random disturbances of latent growth factors. Here we assume that there are no missing observations in $\boldsymbol{x}_i$. The matrix elements of (1.3) are

$$\begin{pmatrix} \eta_{0i} \\ \vdots \\ \eta_{mi} \\ \vdots \\ \eta_{M-1,i} \end{pmatrix} = \begin{pmatrix} \mu_0 \\ \vdots \\ \mu_m \\ \vdots \\ \mu_{M-1} \end{pmatrix} + \begin{pmatrix} \gamma_{01} & \cdots & \gamma_{0k} & \cdots & \gamma_{0,K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mk} & \cdots & \gamma_{m,K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \gamma_{M-1,1} & \cdots & \gamma_{M-1,k} & \cdots & \gamma_{M-1,K} \end{pmatrix} \times \begin{pmatrix} x_{1i} \\ \vdots \\ x_{ki} \\ \vdots \\ x_{Ki} \end{pmatrix} + \begin{pmatrix} \zeta_{0i} \\ \vdots \\ \zeta_{mi} \\ \vdots \\ \zeta_{M-1,i} \end{pmatrix}. \tag{1.4}$$

The measurement model (1.1) or (1.2) denotes the trajectory of the subject's growth over time, while the structural equation model (1.3) or (1.4) represents the subject's differences of their growth factors caused by subject-specific covariates.

1.2 Dropout

Dropout in Longitudinal Studies

Missing values in the analysis of longitudinal study often occur as dropout, which is a special missing pattern in that once participants leave the study,

they do not return. For example, in a study on students' academic achievement, some students may not return to the school during the study period. If the dropout process was directly related to the measurements that could have observed, had he returned to the school, this would become very difficult to handle, and would be beyond the scope of this research.

Terminology

Let us assume without loss of generality, individuals observed up to time T are ordered from those with the complete measurements to those who drop out earliest in the course of study. This forms the monotone pattern shown in Figure 2. For those with complete measurements, $\boldsymbol{y}_{i,obs} = (y_{1i}, \cdots, y_{Ti})^T$ in the data set $\boldsymbol{Y}_{obs} = (\boldsymbol{y}_{1,obs}, \cdots, \boldsymbol{y}_{n,obs})$ is a $T \times 1$ vector of outcomes. Let a $T \times n$ matrix of data $\boldsymbol{Y} = \{\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}\}$, where $\boldsymbol{Y}_{obs}$ consists of the observed part of $\boldsymbol{Y}$ and $\boldsymbol{Y}_{mis}$ represents missing part of $\boldsymbol{Y}$. Let $\boldsymbol{y}_i$ be the ith column of the $T \times n$ matrix $\boldsymbol{Y}$. For each subject i, define that $\boldsymbol{R}_i = (R_{i1}, \cdots, R_{iT})^T$ is $T \times 1$ observation indicator in which $R_{ij} = 1$ if y_{ij} is observed or $R_{ij} = 0$ if y_{ij} is missing. The $\boldsymbol{R}_i = (1, \cdots, 1, 0, \cdots, 0)^T$ means i-th individual drops out from the study at the timepoint at which zero in $\boldsymbol{R}_i$ starts. We let the scalar random variable D_i be the variable indicating at which time dropout occurs for individual i, that is, dropout-time indicator

$$D_i = \sum_{j=1}^{T} R_{ij} + 1.$$

The $D_i = t$ indicates that a subject i drops out between the $(t-1)$-th and t-th observation time; that is, $y_{1i}, \ldots, y_{t-1,i}$ are observed and $y_{ti}, \ldots, y_{Ti}$ are missing. Specially, if $D_i = T + 1$, outcomes for subject i are completely observed.

Types of Dropout

Diggle and Kenward (1994) and Little (1995) developed a classification of dropout process based on the missing value classification framework in Rubin (1976) and Little and Rubin (1987). Let d_i denote the dropout time for subject i, $\boldsymbol{x}_i$ be the design vector for time invariant covariate for subject i, $\boldsymbol{\eta}_i$ be the latent growth factor for $\boldsymbol{y}_i$ for subject i, and $\boldsymbol{\phi}$ be a vector of parameters indexing the dropout-time indicator to be given in (1.6). A strong assumption is that dropout machanism does not depend on outcomes. That is,

$$f(d_i|\boldsymbol{y}_i, \boldsymbol{\phi}) = f(d_i|\boldsymbol{\phi}).$$

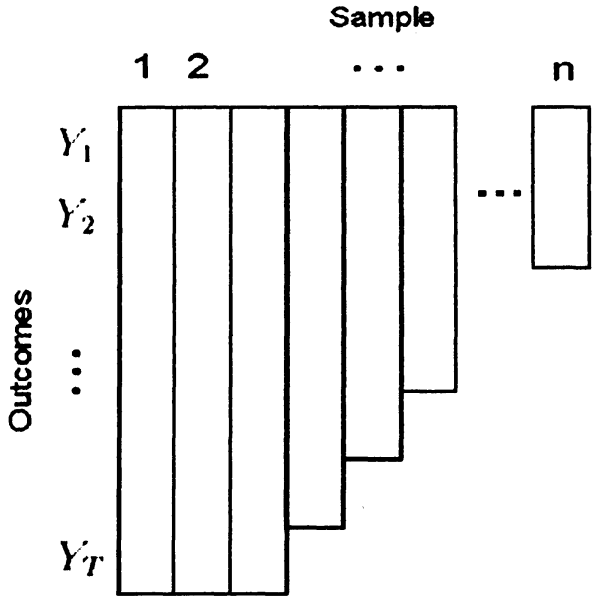

Figure 2: Data set with dropout

Diggle and Kenward (1994) called this assumption "completely random drop-out" (CRD), and viewed it as a special case of MCAR assumption in the original framework.

Suppose that dropout depends only on the observed data, but is independent of the missing outcomes, the distribution of the dropout is

$$f(d_i|\boldsymbol{y}_i, \boldsymbol{\phi}) = f(d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\phi}).$$

Diggle and Kenward (1994) called this condition "random drop-out" (RD), which corresponds to the MAR mechanism.

If dropout depends on missing components of outcome at the time when the subject drops out and possibly the outcomes thereafter, then

$$f(d_i|\boldsymbol{y}_i, \boldsymbol{\phi}) = f(d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{y}_{i,mis}, \boldsymbol{\phi}),$$

where the conditioning on $\boldsymbol{y}_i$ involves components in $\boldsymbol{y}_{i,mis}$. Diggle and Kenward (1994) used the term "informative drop-out" (ID) for this nonignorable mechanism. Under the CRD and RD assumptions, provided that parameters in the measurement process model are independent of the parameters in the dropout process model, we do not have to model dropout mechanism explicitly and separately in measurement process on likelihood-based approach because the parameter estimates are the same as the estimates without missing. Rubin(1976) and Little and Rubin (1987) called this "ignorable." On the other hand, if dropout is not at random, the analytic result without considering dropout process tends to be biased.

Random-Coefficient-Based Dropout

In our dropout modeling, dropout mechanism is supposed to depend on the latent variables (random effects) underlying the observed and missing outcomes. This dropout mechanism assumption appeals to us because influence of outcomes on dropout mechanism is condensed in the latent variables and because if the values of latent variables were inferred, then the dropout mechanism would be determined (Roy and Lin, 2002). Since both the measurement and the dropout processes share random effect in their models, this random effect model for missing or dropout data is called "random-coefficient-based drop-out" in Little (1995), "shared parameter model" in biostatistics, or "sample selection model" in econometrics. Several researcher analyzed their longitudinal model using random-coefficient-based dropout (e.g. Wu and Carroll, 1988; Wu and Bailey, 1989; De Gruttola and Tu, 1994; Follmann and Wu, 1995; Pulkstenis *et al.*, 1998; Ten Have *et al.*, 1998; Alfo and Aitkin, 2000; Ten Have *et al.*, 2000). The crucial condition for the random-coefficient-based dropout is the independence of the measurement process from the dropout process conditional on the random effect underlying those two processes. De Grutolla and Tu (1994), Little (1995), and Ten Have *et al.* (1998) interpreted that the outcomes have no longer information on dropout process model given random effect if random effect is regarded as true response variable measured with error. However if the measurement error is large and random effect does not explain the trajectory of outcomes over time well, then the observed outcome variable may be needed to predict dropout mechanism. Moreover we might be able to interpret that random effect explains dropout tendency over study period while outcome explains temporary trend in dropout process.

In the following we assume a dropout of an individual subject of a longitudinal study depends on his/her past observed measurements and random effects. If so, an advantage of longitudinal studies is that we can partially recover missing information from earlier waves of data and from random effects predicted by covariates in the LGM. Note that in practice, the implementation of random-coefficient-based dropout can be difficult bacause inferences can be quite sensitive to misspecification of the dropout process model (Follumann and Wu, 1995; Ten Have *et al.*, 1998).

Dropout Probability

Diggle and Kenward (1994) took advantage of results in survival analysis as when they formulated their dropout mechanism. We assume that the marginal dropout probability at each time point d_i for subject i is expressed

as a product of two probabilities, the first being the probability that subject i does not drop out up to time $d_i - 1$ expressed as $\prod_{k=2}^{d_i-1}\left\{1 - p(D_i = k|D_i \geq k, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})\right\}$ and the second being the probability that subject i drops out between $d_i - 1$ and d_i expressed as $p(D_i = d_i|D_i \geq d_i, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})$:

$$p(D_i = d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}) = \begin{cases} p(D_i = d_i|D_i \geq d_i, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}) \\ \qquad \times \prod_{k=2}^{d_i-1}\left\{1 - p(D_i = k|D_i \geq k, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})\right\} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } d_i \leq T \\ \prod_{k=2}^{T}\left\{1 - p(D_i = k|D_i \geq k, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})\right\} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } d_i > T \end{cases} \tag{1.5}$$

Consider the conditional probability $p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})$ of dropout at time t for subject i. Since the random variable D_i is categorical, we postulate that the conditional probability is modeled as a logistic linear regression similar to Diggle and Kenward (1994), as

$$\text{logit}\{p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})\} = \phi_0 + \phi_1 y_{t-1,i} + \sum_{j=2}^{M+1} \phi_j \eta_{j-2,i} \equiv \boldsymbol{\phi}^T \boldsymbol{w}_{ti}, \tag{1.6}$$

where $\boldsymbol{\phi} = (\phi_0, \cdots, \phi_{M+1})^T$ is a $(M+2) \times 1$ parameter vector and $\boldsymbol{w}_{ti} = (1, y_{t-1,i}, \eta_{0,i}, \cdots, \eta_{M-1,i})^T$. For brevity, we assume in (1.6) the dropout process depends only on the immediate preceding observed outcome rather than on the previous outcomes. In the matrix notation, the following model for dropout mechanism is thus assumed to hold from (1.6).

$$\log\left(\frac{p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})}{1 - p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})}\right) = \boldsymbol{\phi}^T \boldsymbol{w}_{ti},$$

or

$$\frac{p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})}{1 - p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})} = \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{ti}\right).$$

Since

$$p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}) = \left\{1 - p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})\right\} \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{ti}\right),$$

or

$$p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})\left\{1 + \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{ti}\right)\right\} = \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{ti}\right),$$

we have

$$p(D_i = t|D_i \geq t, \boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}) = \frac{\exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{ti}\right)}{1 + \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{ti}\right)}. \tag{1.7}$$

The dropout probability at time d_i for subject i in (1.5) is thus rewritten by (1.7) as

$$\begin{aligned} p(D_i = d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}) &= \frac{\exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{d_i,i}\right)}{1 + \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{d_i,i}\right)} \times \prod_{k=2}^{d_i-1} \left\{1 - \frac{\exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{k,i}\right)}{1 + \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{k,i}\right)}\right\} \\ &= \frac{\exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{d_i,i}\right)}{1 + \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{d_i,i}\right)} \times \prod_{k=2}^{d_i-1} \frac{1}{1 + \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{k,i}\right)} \\ &= \frac{\exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{d_i,i}\right)}{\prod_{k=2}^{d_i}\left\{1 + \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{k,i}\right)\right\}} \end{aligned} \tag{1.8}$$

if subject i drops out during the study period. Otherwise,

$$p(D_i = d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}) = \frac{1}{\prod_{k=2}^{T}\left\{1 + \exp\left(\boldsymbol{\phi}^T \boldsymbol{w}_{k,i}\right)\right\}}. \tag{1.9}$$

2 Bayesian Estimation of Latent Growth Model with Time Invariant Covariates and with Drop-Out

In this section, we propose our Bayesian approach based on the Lee's Bayesian SEM (Song and Lee, 2002; Lee and Tang, 2006) with the data augmentation and the Markov chain Monte Carlo (MCMC) procedure.

2.1 Likelihood, Priors, and Posteriors

The model parameters and the prior distributions

For subject $i = 1, \ldots, n$, we consider a data set $\boldsymbol{Y}_{obs} = (\boldsymbol{y}_{1,obs}, \ldots, \boldsymbol{y}_{n,obs})$, where $\boldsymbol{y}_{i,obs}$ is a $T \times 1$ vector for subject i with complete observation. A

$T \times n$ matrix $\boldsymbol{Y} = \{\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}\}$ has a monotone pattern of missing data as in Figure 2. We assume that the $\boldsymbol{\epsilon}_i$ in (1.1) and the $\boldsymbol{\zeta}_i$ in (1.3) are independently distributed with respect to i and with respect to each other as

$$\boldsymbol{\epsilon}_i | \boldsymbol{\Psi}_\epsilon \sim MVN(\boldsymbol{0}, \boldsymbol{\Psi}_\epsilon), \tag{2.1}$$

$$\boldsymbol{\zeta}_i | \boldsymbol{\Psi}_\zeta \sim MVN(\boldsymbol{0}, \boldsymbol{\Psi}_\zeta), \tag{2.2}$$

where $\boldsymbol{\Psi}_\epsilon$ is a $T \times T$ diagonal matrix[1] and $\boldsymbol{\Psi}_\zeta$ is a $M \times M$ matrix, not necessarily diagonal, of covariances of $\boldsymbol{\epsilon}_i$'s and $\boldsymbol{\zeta}_i$'s respectively. Suppose $\boldsymbol{\Lambda}$, $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}_\epsilon$, and $\boldsymbol{\Psi}_\zeta$ are model parameters. We need data distribution and the prior distributions to calculate the joint posterior distribution of the parameters. We can specify the distributional form of $\boldsymbol{y}_i$ from (1.1) by incorporating (2.1) as

$$\boldsymbol{y}_i | \boldsymbol{\Lambda}, \boldsymbol{\eta}_i, \boldsymbol{\Psi}_\epsilon \sim MVN(\boldsymbol{\Lambda}\boldsymbol{\eta}_i, \boldsymbol{\Psi}_\epsilon). \tag{2.3}$$

We also know the distribution of $\boldsymbol{\eta}_i$ from (1.3) and (2.2) as

$$\boldsymbol{\eta}_i | \boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{x}_i, \boldsymbol{\Psi}_\zeta \sim MVN(\boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{x}_i, \boldsymbol{\Psi}_\zeta). \tag{2.4}$$

The following conjugate prior distributions will be used to derive the posterior distributions:

$$\boldsymbol{\Lambda}_k | \psi_{\epsilon k} \sim MVN(\boldsymbol{\Lambda}_{0k}, \psi_{\epsilon k}\boldsymbol{H}_{0\epsilon k}), \tag{2.5}$$

$$\psi_{\epsilon k} \sim IG(\alpha_{0\epsilon k}, \beta_{0\epsilon k}), \tag{2.6}$$

$$\boldsymbol{\mu} \sim MVN(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \tag{2.7}$$

$$\widetilde{\boldsymbol{\Gamma}} | \boldsymbol{\Psi}_\zeta \sim MVN(\widetilde{\boldsymbol{\Gamma}}_0, \boldsymbol{\Psi}_\zeta \otimes \boldsymbol{H}_{0\zeta}), \tag{2.8}$$

$$\boldsymbol{\Psi}_\zeta \sim IW(\boldsymbol{R}_0^{-1}, \rho_0), \tag{2.9}$$

where $\boldsymbol{\Lambda}_k^T$ is a $1 \times M$ row vector of unknown parameters in the k-th row of $\boldsymbol{\Lambda}$; $\psi_{\epsilon k}$ is the k-th diagonal element of $\boldsymbol{\Psi}_\epsilon$; $\widetilde{\boldsymbol{\Gamma}}$ is a $MK \times 1$ vector manufactured by the transposed row vectors of $\boldsymbol{\Gamma}$ connected vertically downward, defined as $\mathrm{vec}(\boldsymbol{\Gamma}^T)$ with vec operator; $IG(\alpha_{0\epsilon k}, \beta_{0\epsilon k})$ denotes the inverted Gamma distribution with shape parameter $\alpha_{0\epsilon k}$ and with scale parameter $\beta_{0\epsilon k}$; $IW(\boldsymbol{R}_0^{-1}, \rho_0)$ denotes the inverted Wishart distribution with ρ_0 degrees of freedom and with the precision matrix $\boldsymbol{R}_0^{-1}$; $\boldsymbol{\Lambda}_{0k}, \alpha_{0\epsilon k}, \beta_{0\epsilon k}, \boldsymbol{\mu}_0, \widetilde{\boldsymbol{\Gamma}}_0, \rho_0$, and positive definite matrices $\boldsymbol{H}_{0\epsilon k}, \boldsymbol{H}_{0\zeta}, \boldsymbol{\Sigma}_0, \boldsymbol{R}_0$ are hyper-parameters whose values are assumed to be given by prior information. The $\boldsymbol{\Psi}_\zeta \otimes \boldsymbol{H}_{0\zeta}$ is a $MK \times MK$ matrix denoted by Kronecker product.

[1]This assumption about the disturbance ϵ_i that it is non-autocorrelated, $\mathrm{Cov}(\epsilon_{ti}, \epsilon_{si}) = 0$ $(t \neq s)$, makes (2.14) easy to handle because observed and missing components of $\boldsymbol{y}_i$ become mutually independent.

The likelihood function of the observed data

Let $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\Psi}_\zeta)$ be the parameter vector, $\boldsymbol{\theta}_y = (\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon)$ and $\boldsymbol{\theta}_\eta = (\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\zeta)$ be the parameters included in $\boldsymbol{y}$'s in the measurement model (1.1) and $\boldsymbol{\eta}$'s in the structural equation model (1.3), and $\boldsymbol{\phi}$ be the unknown parameter vector to describe the dropout mechanism. We assume the prior of $\boldsymbol{\phi}$ follows the normal distribution with mean $\boldsymbol{\phi}^0$ and variance $\boldsymbol{V}$, where $\boldsymbol{\phi}^0$ and $\boldsymbol{V}$ are the hyperparameters whose value are assumed to be given by prior information. From (2.3) and (2.4), we can compose the following joint distribution of $\boldsymbol{y}_i$ and $\boldsymbol{\eta}_i$ for individuals $i = 1, \ldots, n$ based on the hierarchical structure.

$$\begin{aligned} & p(\boldsymbol{y}_i|\boldsymbol{\Lambda}, \boldsymbol{\eta}_i, \boldsymbol{\Psi}_\epsilon) \times p(\boldsymbol{\eta}_i|\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\zeta, \boldsymbol{x}_i) \\ &= p(\boldsymbol{y}_i, \boldsymbol{\eta}_i|\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\zeta, \boldsymbol{x}_i) \\ &= p(\boldsymbol{y}_{i,obs}, \boldsymbol{y}_{i,mis}, \boldsymbol{\eta}_i|\boldsymbol{\theta}, \boldsymbol{x}_i). \end{aligned} \tag{2.10}$$

Since the joint distribution of $\boldsymbol{y}_i$ and $\boldsymbol{\eta}_i$, as well as the distribution of dropout time indicator d_i, are indenpendent across individuals i, the joint distribution of the full data $\boldsymbol{Y}$, the latent variable $\boldsymbol{\eta} = \{\boldsymbol{\eta}_i; i = 1, \ldots, n\}$, and the dropout-time indicator $\boldsymbol{D} = \{d_i; i = 1, \ldots, n\}$ can be denoted as

$$\begin{aligned} p(\boldsymbol{Y}, \boldsymbol{\eta}, \boldsymbol{D}|\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\phi}) &= p(\boldsymbol{Y}, \boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{X}) p(\boldsymbol{D}|\boldsymbol{Y}, \boldsymbol{\eta}, \boldsymbol{\phi}) \\ &= \prod_{i=1}^{n} p(\boldsymbol{y}_i, \boldsymbol{\eta}_i|\boldsymbol{\theta}, \boldsymbol{x}_i) p(d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}) \end{aligned} \tag{2.11}$$

under the selection modeling case. The first term on the right hand side in (2.11) is the joint density of the outcome and the latent variable in the latent growth model in (2.10), and the second one is the density of the dropout mechanism, conditional on the observed outcomes and on the latent factors in (1.8) or (1.9). Note that the advantage of the selection model is that it immediately models the distributions which interest us.

Considering that the inference has to be based on observed data, the full data likelihood function is replaced by the observed data likelihood function. The observed data likelihood function $p(\boldsymbol{Y}_{obs}, \boldsymbol{D}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{X})$ can be expressed by averaging the conditional distribution (2.11) over $\boldsymbol{Y}_{mis} = \{\boldsymbol{y}_{i,mis}; i =$

$1, \ldots, n\}$ and $\boldsymbol{\eta}$ as

$$\begin{aligned} p(\boldsymbol{Y}_{obs}, \boldsymbol{D}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{X}) &= \int\int p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{\eta}, \boldsymbol{D}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{\phi})\, d\boldsymbol{\eta} d\boldsymbol{Y}_{mis} \\ &= \int\int p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{X}) p(\boldsymbol{D}|\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{\phi})\, d\boldsymbol{\eta} d\boldsymbol{Y}_{mis} \\ &= \int\int p(\boldsymbol{Y}|\boldsymbol{\eta}, \boldsymbol{\theta}_y) p(\boldsymbol{\eta}|\boldsymbol{\theta}_\eta, \boldsymbol{X}) p(\boldsymbol{D}|\boldsymbol{Y}_{obs}, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{\phi})\, d\boldsymbol{\eta} d\boldsymbol{Y}_{mis}, \end{aligned} \tag{2.12}$$

where the dropout process is referred to as "random-coefficient-based dropout."

The posterior distribution

We assume that individuals i are sampled independently for $i = 1, \ldots, n$. The joint posterior density of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ given $\boldsymbol{Y}_{obs}$ and $\boldsymbol{D}$ is desired and is expressed as

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{D}, \boldsymbol{X}) &\propto p(\boldsymbol{Y}_{obs}, \boldsymbol{D}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{X}) p(\boldsymbol{\theta}, \boldsymbol{\phi}) \\ &= \int\int p(\boldsymbol{Y}|\boldsymbol{\eta}, \boldsymbol{\theta}_y) p(\boldsymbol{\eta}|\boldsymbol{\theta}_\eta, \boldsymbol{X}) p(\boldsymbol{D}|\boldsymbol{Y}_{obs}, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{\phi})\, d\boldsymbol{\eta} d\boldsymbol{Y}_{mis} \\ &\quad \times p(\boldsymbol{\theta}, \boldsymbol{\phi}) \\ &= \int\int p(\boldsymbol{Y}|\boldsymbol{\eta}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon) p(\boldsymbol{\eta}|\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\zeta, \boldsymbol{X}) p(\boldsymbol{D}|\boldsymbol{Y}_{obs}, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{\phi}) \\ &\quad \times \prod_{k=1}^{T} \{p(\boldsymbol{\Lambda}_k|\psi_{\epsilon k}) p(\psi_{\epsilon k})\} p(\boldsymbol{\mu}) p(\widetilde{\boldsymbol{\Gamma}}|\boldsymbol{\Psi}_\zeta) p(\boldsymbol{\Psi}_\zeta) p(\boldsymbol{\phi})\, d\boldsymbol{\eta} d\boldsymbol{Y}_{mis}, \end{aligned} \tag{2.13}$$

under the selection modeling case. The observed data likelyhood function is decomposed into the density of the outcome variable, the density of the latent growth variable, and the conditional density of the latent factor given the covariates, the density of the dropout mechanism, conditional on the observed outcomes and on the latent growth factors. Note that the advantage of the selection model is that it immediately models the distributions which interest us.

Owing to the existence of missing data $\boldsymbol{Y}_{mis}$ and the existence of the latent variable $\boldsymbol{\eta}$, the likelihood function (2.12) has multiple integrals. Thus the posterior distribution (2.13) is difficult to calculate. Taking advantage of current statistical computing, Song and Lee (2002) and Lee and Tang (2006) conducted posterior analyses with the data augmentation technique

and the MCMC algorithm. They introduced the missing data set $\boldsymbol{Y}_{mis}$ to the posterior distribution as unknown parameter to estimate. In addition to "real" missing data, they treated the latent variables $\boldsymbol{\eta}$ as a hypothetical missing data. We call $(\boldsymbol{\eta}, \boldsymbol{Y}_{mis})$ latent data.

The distribution of the missing data

Now we introduce the distribution of the missing data under random-coefficient-based dropout, which is required in the data augmentation and the MCMC. Note that $\boldsymbol{y}_i = (\boldsymbol{y}_{i,obs}^T, \boldsymbol{y}_{i,mis}^T)^T$, where $\boldsymbol{y}_{i,obs}$ is the set of observed data of $\boldsymbol{y}_i$ with $d_i - 1$ elements and $\boldsymbol{y}_{i,mis}$ is the set of missing components consisting of NAs of $\boldsymbol{y}_i$ with $T - d_i + 1$ elements.

The component expression of measurement model with dropout for subject i is denoted from (1.1) as

$$\begin{pmatrix} y_{1i} \\ \vdots \\ y_{d_i-1,i} \\ [y_{i,mis}^{d_i}] \\ \vdots \\ [y_{i,mis}^{T}] \end{pmatrix} = \begin{pmatrix} \lambda_{10} & \cdots & \lambda_{1,M-1} \\ \vdots & \ddots & \vdots \\ \lambda_{d_i-1,0} & \cdots & \lambda_{d_i-1,M-1} \\ [\lambda_{i,mis}^{d_i,0} & \cdots & \lambda_{i,mis}^{d_i,M-1}] \\ \vdots & \ddots & \vdots \\ [\lambda_{i,mis}^{T0} & \cdots & \lambda_{i,mis}^{T,M-1}] \end{pmatrix} \begin{pmatrix} \eta_{0i} \\ \vdots \\ \eta_{M-1,i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \vdots \\ \epsilon_{d_i-1,i} \\ [\epsilon_{i,mis}^{d_i}] \\ \vdots \\ [\epsilon_{i,mis}^{T}] \end{pmatrix},$$

where the notation $[\cdot]$ denotes missing component and the upper-right indices denote the location of these elements within their corresponding vectors or matrices. The $\boldsymbol{y}_i$'s are normally distributed and $\boldsymbol{\Psi}_\epsilon$ is a diagonal matrix, so that $\boldsymbol{y}_{i,mis}$ given $(\boldsymbol{\theta}_y, \boldsymbol{\eta}_i)$ is independent with $\boldsymbol{y}_{i,obs}$. The random-coefficient-based dropout assumption and the non-autocorrelation assumption imply that the missing data in $\boldsymbol{y}_i$ directly depend only on the latent variable $\boldsymbol{\eta}_i$ not $\boldsymbol{y}_i$ itself.

For $i = 1, \ldots, n$, since $\boldsymbol{y}_i|\boldsymbol{\theta}, \boldsymbol{\eta}_i$ are assumed independent with respect to i, $\boldsymbol{y}_{i,mis}$'s given $(\boldsymbol{\theta}, \boldsymbol{\eta}_i)$'s are also independent. Therefore the full conditional distribution of $\boldsymbol{Y}_{mis}$ given $\boldsymbol{\theta}$, $\boldsymbol{\eta}$, $\boldsymbol{Y}_{obs}$, $\boldsymbol{X}$, $\boldsymbol{D}$, $\boldsymbol{\phi}$ for the Gibbs sampler in the MCMC is

$$\begin{aligned} p(\boldsymbol{Y}_{mis}|\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}, \boldsymbol{\phi}) &= \prod_{i=1}^{n} p(\boldsymbol{y}_{i,mis}|\boldsymbol{\eta}_i, \boldsymbol{\theta}, \boldsymbol{y}_{i,obs}, \boldsymbol{x}_i, d_i, \boldsymbol{\phi}) \\ &= \prod_{i=1}^{n} p(\boldsymbol{y}_{i,mis}|\boldsymbol{\eta}_i, \boldsymbol{\theta}_y) \end{aligned} \tag{2.14}$$

and the right hand side of this equation is

$$
\begin{aligned}
p(\boldsymbol{y}_{i,mis}|\boldsymbol{\eta}_i, \boldsymbol{\theta}_y) &= |\boldsymbol{\Psi}_{\epsilon,i,mis}|^{-\frac{1}{2}} \\
&\quad \times \exp\left\{ -\frac{1}{2}(\boldsymbol{y}_{i,mis} - \boldsymbol{\Lambda}_{i,mis}\boldsymbol{\eta}_i)^T \boldsymbol{\Psi}^{-1}_{\epsilon,i,mis}(\boldsymbol{y}_{i,mis} - \boldsymbol{\Lambda}_{i,mis}\boldsymbol{\eta}_i) \right\} \\
&= \prod_{k=1}^{T-d_i+1} |\psi_{\epsilon,i,mis,k}|^{-\frac{1}{2}} \\
&\quad \times \exp\left\{ -\frac{1}{2}(y_{i,mis,k} - \boldsymbol{\Lambda}_{i,mis,k}\boldsymbol{\eta}_i)^T \psi^{-1}_{\epsilon,i,mis,k}(y_{i,mis,k} - \boldsymbol{\Lambda}_{i,mis,k}\boldsymbol{\eta}_i) \right\},
\end{aligned}
\tag{2.15}
$$

where the $\boldsymbol{\Lambda}_{i,mis}$ is a $(T - d_i + 1) \times M$ submatrix of $\boldsymbol{\Lambda}$ with rows corresponding to missing components and $\boldsymbol{\Psi}_{\epsilon,i,mis}$ is a $(T - d_i + 1) \times (T - d_i + 1)$ submatrix of $\boldsymbol{\Psi}_\epsilon$ with rows and columns corresponding to missing values. The index k denotes the k-th element, diagonal element, or row vector in corresponding vector or matrices. As shown in (2.15), even the form of $\boldsymbol{Y}_{mis}$ is complicated with any monotone pattern of dropout, its conditional distribution only involves a product of normal distributions. As a result, the computational burden for simulating $\boldsymbol{Y}_{mis}$ is light. In this sense, the Bayesian estimating procedure under the random-coefficient-based dropout and the non-autocorrelated measurement error is almost the same as the parameter estimation procedure under the complete data set with the data augmentation and the MCMC.

2.2 Data Augmentation and MCMC Procedure

The data augmentation

We form an algorithm of the data augmentation technique for the desired joint posterior density $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ of $(\boldsymbol{\theta}, \boldsymbol{\phi})$ given the observed data and the dropout indicator as follows.

$$
\begin{aligned}
&p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) \\
&\quad = \int\int p(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})\, d\boldsymbol{\eta} d\boldsymbol{Y}_{mis} \\
&\quad = \int\int p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{\eta}, \boldsymbol{Y}_{mis}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})\, d\boldsymbol{\eta} d\boldsymbol{Y}_{mis},
\end{aligned}
\tag{2.16}
$$

where $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{\eta}, \boldsymbol{Y}_{mis}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ denotes the joint conditional density of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ given the latent data augmented as $(\boldsymbol{\eta}, \boldsymbol{Y}_{mis})$ and the observed

variables $\boldsymbol{Y}_{obs}$, $\boldsymbol{X}$, and $\boldsymbol{D}$ and $p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ denotes the joint predictive density of the latent data $(\boldsymbol{\eta}, \boldsymbol{Y}_{mis})$ given $\boldsymbol{Y}_{obs}$, $\boldsymbol{X}$, and $\boldsymbol{D}$. The joint predictive density $p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ of the latent data $(\boldsymbol{\eta}, \boldsymbol{Y}_{mis})$ relates to the desired posterior density $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ as

$$p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) = \int\int p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) d\boldsymbol{\theta} d\boldsymbol{\phi}. \tag{2.17}$$

By substituting (2.17) into (2.16), we can in principle form an iterative algorithm to obtain the desired posterior density.

$$\begin{aligned} & p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) \\ & = \int\int p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{\eta}, \boldsymbol{Y}_{mis}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})\, d\boldsymbol{\eta} d\boldsymbol{Y}_{mis} \\ & = \int\int p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{\eta}, \boldsymbol{Y}_{mis}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) \\ & \quad \times \left\{ \int\int p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{Y}_{\hat{o}bs}, \boldsymbol{X}, \boldsymbol{D}) p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})\, d\boldsymbol{\theta} d\boldsymbol{\phi} \right\} d\boldsymbol{\eta} d\boldsymbol{Y}_{mis}. \end{aligned} \tag{2.18}$$

The fact that $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ appears on both sides of (2.18) gives the following iterative algorithm. Given the values of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, we generate $(\boldsymbol{\eta}_l, \boldsymbol{Y}_{mis,l})$ for $l = 1, \ldots, L$ from the joint density $p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ of the latent data $(\boldsymbol{\eta}, \boldsymbol{Y}_{mis})$ given $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, $\boldsymbol{Y}_{obs}$, $\boldsymbol{X}$, and $\boldsymbol{D}$. This method is called the composition method, which reexpresses $p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ as the expression in the brace in (2.18) so that we can obtain random draws of $(\boldsymbol{\eta}, \boldsymbol{Y}_{mis})$ from $p(\boldsymbol{\eta}, \boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$. With these samples $(\boldsymbol{\eta}_l, \boldsymbol{Y}_{mis,l})$, we manufacture $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{\eta}_l, \boldsymbol{Y}_{mis,l}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ and approximate the desired posterior density $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ by

$$g(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}) = \frac{1}{L} \sum_{l=1}^{L} p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{\eta}_l, \boldsymbol{Y}_{mis,l}, \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D}). \tag{2.19}$$

Assume the $g(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ is a good approximation of the posterior $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$, we generate $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ and we repeat this process until convergence by the standard Monte Carlo.

The MCMC procedure

We propose the following algorithm. Note that we use the Metropolis-Hastings algorithm for the non-standard conditional distributions.

MCMC 0 Set $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\phi}^{(0)}$, and $\boldsymbol{Y}_{mis}^{(0)}$.

At the j-th iteration $j = 1, \ldots,$

MCMC 1 For $i = 1, \ldots, n$,
generate $\boldsymbol{\eta}_i^*$ from $MVN(\boldsymbol{B}_\eta \boldsymbol{b}_\eta, \boldsymbol{B}_\eta)$ with $\boldsymbol{B}_\eta = (\boldsymbol{\Psi}_\zeta^{-1} + \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1} \boldsymbol{\Lambda})^{-1}$ and $\boldsymbol{b}_\eta = \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1} \boldsymbol{y}_i + \boldsymbol{\Psi}_\zeta^{-1}(\boldsymbol{\mu} + \boldsymbol{\Gamma} \boldsymbol{x}_i)$.

MCMC 2 Calculate

$$R_{\eta_i^*} = \min\left(1, \frac{p(d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i^*, \boldsymbol{\phi})}{p(d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i^{(j-1)}, \boldsymbol{\phi})}\right)$$

MCMC 3 Set $\boldsymbol{\eta}_i^{(j)} = \boldsymbol{\eta}_i^*$ with probability $R_{\eta_i^*}$ or $\boldsymbol{\eta}_i^{(j)} = \boldsymbol{\eta}_i^{(j-1)}$ with probability $1 - R_{\eta_i^*}$.

MCMC 4 For $i = 1, \ldots, n$,
generate $\boldsymbol{y}_{i,mis}^{(j)}$ from $MVN(\boldsymbol{\Lambda}_{i,mis} \boldsymbol{\eta}_i, \boldsymbol{\Psi}_{\epsilon,i,mis})$ in (2.15).

MCMC 5 For $k = 1, \ldots, T$,
generate $\boldsymbol{\Lambda}_k^{(j)}$ from $MVN(\boldsymbol{a}_k, \boldsymbol{A}_k)$ with $\boldsymbol{A}_k = (\boldsymbol{H}_{0\epsilon k}^{-1} + \boldsymbol{\eta}\boldsymbol{\eta}^T)^{-1}$ and $\boldsymbol{a}_k = \boldsymbol{A}_k(\boldsymbol{H}_{0\epsilon k}^{-1} \boldsymbol{\Lambda}_{0k} + \boldsymbol{\eta} \boldsymbol{Y}_k)$.

MCMC 6 For $k = 1, \ldots, T$,
generate $\psi_{\epsilon k}^{(j)}$ from $IG\Big(\frac{n}{2} + \alpha_{0\epsilon k}, \beta_{0\epsilon k} + 2^{-1}\big(\boldsymbol{Y}_k^T \boldsymbol{Y}_k - \boldsymbol{a}_k^T \boldsymbol{A}_k^{-1} \boldsymbol{a}_k + \boldsymbol{\Lambda}_{0k}^T \boldsymbol{H}_{0\epsilon k}^{-1} \boldsymbol{\Lambda}_{0k}\big)\Big)$.

MCMC 7 Generate $\boldsymbol{\mu}^{(j)}$ from $MVN(\boldsymbol{b}_\mu, \boldsymbol{B}_\mu)$ with $\boldsymbol{B}_\mu = (\boldsymbol{\Sigma}_0^{-1} + n\boldsymbol{\Psi}_\zeta^{-1})^{-1}$ and $\boldsymbol{b}_\mu = n\boldsymbol{\Psi}_\zeta^{-1} \bar{\boldsymbol{\eta}} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0$.

MCMC 8 Generate $\widetilde{\boldsymbol{\Gamma}}^{(j)}$ from $MVN(\widetilde{\boldsymbol{a}}_\Gamma, \boldsymbol{\Psi}_\zeta \otimes \boldsymbol{A}_\Gamma)$ with $\widetilde{\boldsymbol{a}}_\Gamma = \mathrm{vec}(\boldsymbol{a}_\Gamma^T)$, $\boldsymbol{a}_\Gamma^T = \boldsymbol{A}_\Gamma\{\boldsymbol{H}_{0\zeta}^{-1} \boldsymbol{\Gamma}_0^T + \boldsymbol{X}(\boldsymbol{\eta} - \boldsymbol{\mu})^T\}$, and $\boldsymbol{A}_\Gamma = (\boldsymbol{H}_{0\zeta}^{-1} + \boldsymbol{X}\boldsymbol{X}^T)^{-1}$.

MCMC 9 Generate $\boldsymbol{\Psi}_\zeta^{(j)}$ from $IW(\boldsymbol{R}_0^{-1} + \boldsymbol{S}, \rho_0 + n)$ with $\boldsymbol{S} = (\boldsymbol{\eta} - \boldsymbol{\mu} - \boldsymbol{a}_\Gamma \boldsymbol{X})(\boldsymbol{\eta} - \boldsymbol{\mu} - \boldsymbol{a}_\Gamma \boldsymbol{X})^T + (\boldsymbol{a}_\Gamma - \boldsymbol{\Gamma}_0)\boldsymbol{H}_{0\zeta}^{-1}(\boldsymbol{a}_\Gamma - \boldsymbol{\Gamma}_0)^T$.

MCMC 10 Generate $\boldsymbol{\phi}^*$ from $MVN(\boldsymbol{\phi}^{(j-1)}, \sigma_\phi^2 \boldsymbol{\Omega}_\phi)$ with $\boldsymbol{\Omega}_\phi^{-1} = \boldsymbol{V}^{-1} + \sum_{i=1}^n \sum_{k=2}^{d_i} w_{ki} w_{ki}^T$.

MCMC 11 Calculate

$$R_{\phi^*} = \min\left(1, \prod_{i=1}^n \frac{p(d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}^*)p(\boldsymbol{\phi}^*)}{p(d_i|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi}^{(j-1)})p(\boldsymbol{\phi}^{(j-1)})}\right)$$

MCMC 12 Set $\boldsymbol{\phi}^{(j)} = \boldsymbol{\phi}^*$ with probability R_{ϕ^*} or $\boldsymbol{\phi}^{(j)} = \boldsymbol{\phi}^{(j-1)}$ with probability $1 - R_{\phi^*}$.

MCMC 13 If the conditional distribution $p(\boldsymbol{\theta}_y|\boldsymbol{\eta}, \boldsymbol{Y})$ converges when using the conditional distribution in **MCMC 5** and **MCMC 6**, if the conditional distribution $p(\boldsymbol{\theta}_\eta|\boldsymbol{\eta}, \boldsymbol{X})$ converges as well when using the conditional distribution in **MCMC 7** to **MCMC 9**, and if the conditional distribution $p(\boldsymbol{\phi}|\boldsymbol{\eta}, \boldsymbol{Y}_{obs}, \boldsymbol{D})$ converges in **MCMC 10**, then $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{D})$ is the stationary distribution. Hence, stop the iteration. Otherwise return to **MCMC 1**.

3 Simulation Study

We will examine the accuracy of the proposed estimates. The complete data set $\boldsymbol{Y}$ with 100 subjects and 5 time points are generated 100 times from the model in (1.1) and (1.3). The values of the population parameters are set as below.

$$
\begin{aligned}
\boldsymbol{\Lambda}^T &= \begin{pmatrix} 1.0^* & 1.0^* & 1.0^* & 1.0^* & 1.0^* \\ 0.0^* & 1.0^* & 2.0 & 3.0 & 4.0 \end{pmatrix}, \\
\boldsymbol{\Psi}_\epsilon &= \mathrm{diag}(1.0, \cdots, 1.0), \\
\boldsymbol{\mu}^T &= (0.0, 0.0), \\
\boldsymbol{\Gamma} &= \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}, \\
\boldsymbol{\Psi}_\zeta &= \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{pmatrix}, \\
\boldsymbol{\phi}^T &= (-1.0, 0.5, 0.5, 0.5),
\end{aligned}
$$

The asterisks in $\boldsymbol{\Lambda}^T$ are the values fixed for identification and will not be estimated. For subject i, we generate randomly $\boldsymbol{x}_i^T = (x_{1i}, x_{2i})$ and $\boldsymbol{\zeta}_i^T = (\zeta_{0i}, \zeta_{1i})$ from $MVN(\boldsymbol{0}, \boldsymbol{I})$ and $MVN(\boldsymbol{0}, \boldsymbol{\Psi}_\zeta)$ respectively, and calculate $\boldsymbol{\eta}_i$ in (1.3). Then we calculate $\boldsymbol{y}_i$ in (1.1) by using random $\boldsymbol{\epsilon}_i$ generated from $MVN(\boldsymbol{0}, \boldsymbol{\Psi}_\epsilon)$ and the $\boldsymbol{\eta}_i$. Repeating these steps for 100 subjects, we obtain a complete set of the simulated data. Missing data satisfying random-coefficient-based dropout mechanism in (1.8) are created via the following steps:

(i) We select 75 out of 100 subjects randomly as possible candidates for dropouts. This reduces the number of subjects with dropout to be less than or equal to 75.

(ii) We generate a random number ν from the uniform distribution $U(0,1)$. This will be used for all the subjects and for all the simulated sets.

(iii) For each of the selected 75 subjects, if $\nu \leq p(D_i = t|\boldsymbol{y}_{i,obs}, \boldsymbol{\eta}_i, \boldsymbol{\phi})$ in (1.8) for $t = 2, \ldots, 5$, then we change $y_{ti}, \ldots, y_{5i}$ to be missing.

We repeat (i) and (iii) for 100 times to generate 100 sets of the simulated data with dropouts. After taking these steps, we have roughly 43 subjects dropping out on average for the 100 simulated data with dropouts. See Figure 3. In Figure 4 we present the overall picture as to when the dropouts occur in the 100 simulated data sets, each of which consist of 100 subjects. Of the total 4,341 subjects who drop out sometime during the five observation sequence, 3,396 drop out at time 2, 654 at time 3, 202 at time 4, and 89 at time 5. Since the remaining 5,659 subjects are fully observed, of the $5 \times 100 \times 100$ data points, the average proportion of $4 \times 3{,}396 + 3 \times 654 + 2 \times 202 + 1 \times 89 = 14{,}734$ or 29.47% are missing.

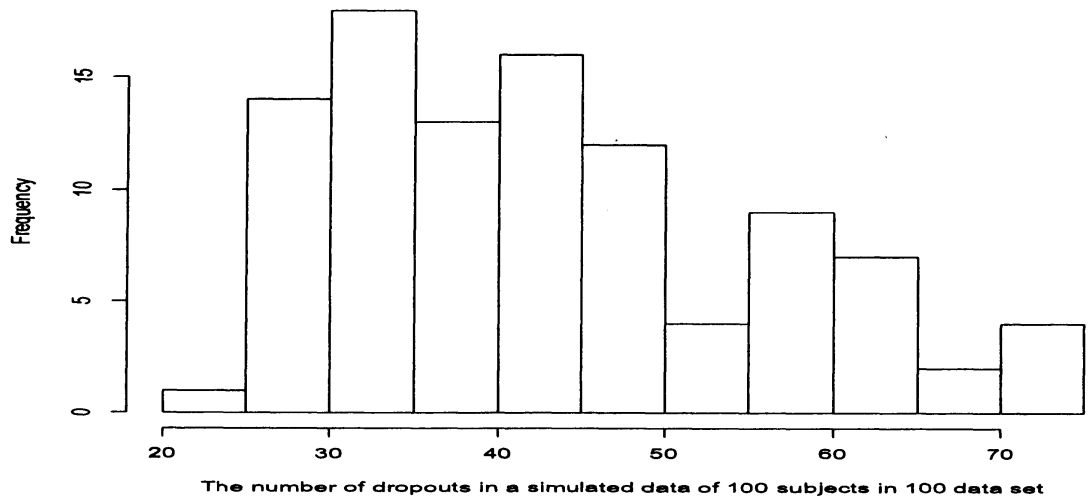

Figure 3: The histogram of the number of dropouts in a simulated data of 100 subjects in 100 data sets.

We obtain the estimates using three types of data:

A 100 sets of the simulated complete data.

B 100 sets of the simulated data with dropouts.

C 100 sets of the simulated but listwisely deleted data when subjects drop out sometime during the five observation sequence. Because of listwise deletion, the number of subjects in a set may vary from 26 to 76.

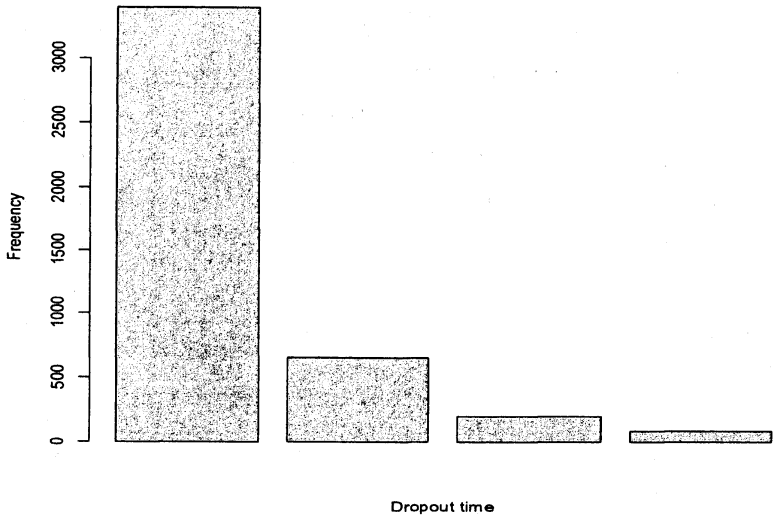

Figure 4: The total number of subjects dropping out at each time

For the result in Table 1, three data sets are estimated using the Bayesian framework. For the data set A and C, there are no dropouts and so the estimation is done using the proposed algorithm in principle, but those parts dealing with the dropouts are removed.

The hyper-parameter values in priors are set as $\lambda_{0k} = k - 1$ for $k = 3, \ldots, 5$, $\alpha_{0\epsilon k} = 10$, $\beta_{0\epsilon k} = 8$, $\boldsymbol{\mu}_0 = (0,0)^T$, $\widetilde{\boldsymbol{\Gamma}}_0 = (1.0, \cdots, 1.0)^T$, $\rho_0 = 8$, $\boldsymbol{\phi}^0 = (-0.5, 0.25, 0.25, 0.25)^T$, a location parameter $c = 0.25$ for ψ_ϵ's, $\boldsymbol{\Sigma}_0$ is diagonal matrix with diagonally element 0.01, $\boldsymbol{H}_{0\zeta}$ and $\boldsymbol{V}$ are diagonal matrices with 0.25, and $\boldsymbol{R}_0^{-1}$ is 5 times indentity matrix. The variance σ_ϕ^2 in the M-H algorithm is chosen as 0.05 to give acceptance probability almost 0.4. For the data sets A and C, $\boldsymbol{\phi}^0$, $\boldsymbol{V}$, and σ_ϕ^2 are of course unnecessary. This setting can be regarded as a situation under good prior information.

We conduct the Bayesian estimation based on 5,000 iterations after throwing out the first 5,000 burn-in. The posterior means and their standard errors (SE) for the parameters are computed. Results are given in Table 1. We see that the estimates λ's, μ's, γ's, ψ_ϵ's, and ψ_ζ's under our proposed method in the fourth column from the left are similar to those under the complete data in the second column from the left, assuming we know the true parameter values. In the dropout process model, ϕ_2 and ϕ_3 are overestimated while ϕ_0 and ϕ_1 are underestimated relative to the corresponding true values. In one of the preceding studies, Roy and Lin (2002) observed similar compensated tendency of parameter estimates in dropout process in their sensitivity analysis when they formulate the dropout probability in terms of logistic regression, although they specified that the dropout probabilities depend on

both the last observed and the current missing outcomes. Note that our dropout probabilities depend on the latent variables as well. The standard errors of λ's, μ's, γ's, ψ_ϵ's, and ψ_ζ's in the fifth column from the left under the proposed model is larger relative to those in the third column under the complete data. This elevated standard errors are observed for the following two reasons: First, under the proposed model, there are four more parameters to be estimated with the same number of data points, thereby increasing the fluctuation of the parameter estimates; Second, by introducing the dropout process, we also introduce the number of data available for estimation to be variable as well as seen in Figure 3. Table 2 shows the average of the 100 maximum likelihood (ML) estimates and their corresponding standard errors using the software **Mplus** 4.21 (Muthén & Muthén, 2006) for the data set A and C. The ML estimates resemble our Bayesian estimates in the mean estimates. Notice that the underestimation of the estimate for $\psi_{\zeta 12}$ in any one of the five cases. Hence we conclude that the severe underestimation observed in the fourth column of Table 1 is not due to the algorithm, but because of the unfortunate chance consequence of the simulated data.

Reference

Alfo, M. & Aitkin, M. (2000). Random coefficient models for binary longitudinal responses with attrition. *Statistics and Computing*, 10, 279-287.

De Gruttola, V. & Tu, X. M. (1994). Modelling progression of CD4-Lymphocyte count and its relationship to survival time. *Biometrics*, 50(4), 1003-1014.

Diggle, P. & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics*, 43(1), 49-93.

Follmann, D. & Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 51(1), 151-168.

Lee, S. Y. & Tang, N. S. (2006). Bayesian analysis of nonlinear structural equation models with nonignorable missing data. *Psychometrika*, 71(3), 541-564.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of American Statistical Association*, 90, 431, 1112-1121.

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

Pulkstenis, E. P., Ten Have, T. R., & Richard Landis, J. (1998). Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*, 93, 442, 438-450.

Table 1: Performance of the Bayesian estimates to examine the accuracy

True parameter values	Complete (A)		Dropout (B)		Lst Delete (C)	
	Mean	SE	Mean	SE	Mean	SE
$\lambda_{32} = 2.0$	2.017	0.096	2.032	0.136	2.022	0.130
$\lambda_{42} = 3.0$	3.019	0.137	3.031	0.191	3.027	0.170
$\lambda_{52} = 4.0$	4.024	0.172	4.042	0.256	4.048	0.207
$\mu_1 = 0.0$	0.010	0.077	-0.051	0.088	-0.069	0.051
$\mu_2 = 0.0$	-0.004	0.062	0.107	0.146	-0.068	0.054
$\gamma_{11} = 1.0$	0.990	0.115	0.974	0.125	1.003	0.144
$\gamma_{12} = 0.5$	0.482	0.110	0.480	0.115	0.448	0.147
$\gamma_{21} = 0.5$	0.531	0.088	0.649	0.163	0.554	0.127
$\gamma_{22} = 1.0$	0.968	0.095	1.054	0.221	0.917	0.113
$\psi_{\epsilon 1} = 1.0$	1.020	0.179	1.074	0.280	0.967	0.157
$\psi_{\epsilon 2} = 1.0$	0.991	0.149	1.012	0.219	0.951	0.165
$\psi_{\epsilon 3} = 1.0$	0.957	0.134	1.057	0.465	0.924	0.158
$\psi_{\epsilon 4} = 1.0$	0.983	0.147	1.225	0.890	0.964	0.161
$\psi_{\epsilon 5} = 1.0$	0.939	0.160	1.427	1.661	0.970	0.172
$\psi_{\zeta 11} = 1.0$	0.957	0.211	1.007	0.251	0.939	0.248
$\psi_{\zeta 12} = 0.3$	0.044	0.121	0.168	0.231	0.089	0.146
$\psi_{\zeta 22} = 1.0$	1.011	0.166	1.603	0.871	0.976	0.185
$\phi_0 = -1.0$	-	-	-1.879	0.557	-	-
$\phi_1 = 0.5$	-	-	0.131	0.610	-	-
$\phi_2 = 0.5$	-	-	0.719	0.856	-	-
$\phi_3 = 0.5$	-	-	0.789	0.696	-	-

Table 2: Performance of the maximum likelihood estimates using Mplus

True parameter values	Complete (A)		Lst Delete (C)	
	Mean	SE	Mean	SE
$\lambda_{32} = 2.0$	2.000	0.016	2.037	0.021
$\lambda_{42} = 3.0$	2.987	0.024	3.044	0.032
$\lambda_{52} = 4.0$	3.984	0.033	4.069	0.044
$\mu_1 = 0.0$	0.007	0.013	-0.051	0.088
$\mu_2 = 0.0$	-0.011	0.011	0.107	0.146
$\gamma_{11} = 1.0$	0.994	0.013	0.929	0.017
$\gamma_{12} = 0.5$	0.497	0.012	0.425	0.016
$\gamma_{21} = 0.5$	0.511	0.012	0.453	0.015
$\gamma_{22} = 1.0$	1.009	0.013	0.919	0.016
$\psi_{\epsilon 1} = 1.0$	1.018	0.024	0.947	0.030
$\psi_{\epsilon 2} = 1.0$	0.999	0.018	0.951	0.023
$\psi_{\epsilon 3} = 1.0$	0.987	0.017	0.946	0.022
$\psi_{\epsilon 4} = 1.0$	1.034	0.023	1.018	0.030
$\psi_{\epsilon 5} = 1.0$	0.938	0.036	1.007	0.048
$\psi_{\zeta 11} = 1.0$	0.999	0.024	0.974	0.031
$\psi_{\zeta 12} = 0.3$	0.002	0.014	0.006	0.018
$\psi_{\zeta 22} = 1.0$	0.997	0.024	0.880	0.028

Roy, J & Lin, X. (2002). Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: Changes in methadone treatment practices. *Journal of the American Statistical Association*, 97, 457, 40-52.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

Song, X. Y. & Lee, S. Y. (2002). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika*, 67(2), 261-288.

Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., & Richard Landis, J. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, 54(1), 367-383.

Ten Have, T. R., Miller, M. E., Reboussin, B. A., & James, M. K. (2000). Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the longitudinal study of aging. *Biometrics*, 56(1), 279-287.

Wu, M. C. & Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics*, 45(3), 939-955.

Wu, M. C. & Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1), 175-188.