Published in final edited form as:
Biometrics. 2017 March ; 73(1): 313–323. doi:10.1111/biom.12536.

Hidden Markov Latent Variable Models with Multivariate Longitudinal Data

Xinyuan Song[1,*], **Yemao Xia**[2], and **Hongtu Zhu**[3,**]

[1]Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

[2]Department of Applied Mathematics, Nanjing Forestry University, Nanjing, China

[3]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

Summary

Cocaine addiction is chronic and persistent, and has become a major social and health problem in many countries. Existing studies have shown that cocaine addicts often undergo episodic periods of addiction to, moderate dependence on, or swearing off cocaine. Given its reversible feature, cocaine use can be formulated as a stochastic process that transits from one state to another, while the impacts of various factors, such as treatment received and individuals' psychological problems on cocaine use, may vary across states. This paper develops a hidden Markov latent variable model to study multivariate longitudinal data concerning cocaine use from a California Civil Addict Program. The proposed model generalizes conventional latent variable models to allow bidirectional transition between cocaine-addiction states and conventional hidden Markov models to allow latent variables and their dynamic interrelationship. We develop a maximum likelihood approach, along with a Monte Carlo expectation conditional maximization (MCECM) algorithm, to conduct parameter estimation. The asymptotic properties of the parameter estimates and statistics for testing the heterogeneity of model parameters are investigated. The finite sample performance of the proposed methodology is demonstrated by simulation studies. The application to cocaine use study provides insights into the prevention of cocaine use.

Keywords

Hidden Markov model; latent variables; log-continuation ratio model; MCECM algorithm; multivariate longitudinal data

1. Introduction

We consider a longitudinal study on cocaine use carried out by the UCLA center for advancing longitudinal drug abuse research. In this study, 321 participants admitted in 1988–89 to the West Los Angeles Veterans Affairs Medical Center were assessed at baseline, one

---

[*] xysong@sta.cuhk.edu.hk. [**] htzhu@email.unc.edu.

Supplementary Materials
Web Appendices referenced in Sections 3 and 4, and the computer code are available with this paper at the *Biometrics* website on Wiley Online Library.

year after treatment, two years after treatment, and 12 years after treatment in 2002–03. Interview questionnaire covers information on the participants' cocaine use behavior, treatment received, and psychological problems. Cocaine use behavior is fully measured by one of the questionnaire items, whereas treatment and psychological problems are summarized by two or more questionnaire items and are therefore regarded as latent traits. A primary interest of this study is to investigate the effects of latent traits, such as treatment and psychological problems on cocaine use behavior. We propose the use of latent variable model (LVM) to examine the interrelationships between the observed and latent variables. Moreover, unlike an irreversible and progressive event, cocaine use process often comprises episodic periods of addiction to, moderate dependence on, and swearing off cocaine. Thus, identifying latent states from continuous cocaine use and investigating its transition pattern is also of interest. The aim of this paper is to develop a hidden Markov model (HMM) to characterize the temporal latent process of cocaine use along with its latent risk factors and the bidirectional transition between various cocaine addiction states.

However, most existing HMMs in the literature cannot adequately address three major dependency structures in multivariate longitudinal data, including the correlation among multiple responses within the same subject, temporal dependence, and heterogeneity. See Cappé, Moulines and Rydén (2005) for a comprehensive review of HMMs. A basic assumption of HMMs is that the latent discrete process is a first-order Markov chain, and that occasion-specific response variables can be modeled as an independent process conditioning on the sequence of latent states. An initial approach has been developed by Vermunt et al. (1999) and further developed, in the context of multivariate data, by many researchers. Scott, James, and Sugar (2005) described an HMM for continuous and multivariate t-distributed data, whereas Altman (2007) proposed mixed-effects HMMs and identified a two-state Poisson model for lesion-count data. Bartolucci and Farcomeni (2009) and Bartolucci et al. (2009) developed dynamic logit models for analyzing longitudinal categorical data and investigated the conditional probability of categorical response across time. Maruotti (2011) further analyzed longitudinal binary and count data using a mixed HMM within the generalized linear random effect model framework, wherein the conditional model incorporates random effects but the transition model neither includes random effects nor depends on occasions and subjects (see the book of Bartolucci et al. (2013) for an overview). Recently, Chow et al. (2013) utilized a multinomial logistic regression model to characterize bidirectional transitions between latent classes as well as a LVM to identify the class-specific association structure among observed and latent variables. However, their transition model failed to incorporate random effects and the order of hidden states, and is therefore restricted to the modeling of Markovian process and incapable of revealing the heterogeneity of transition process as well as the ordered feature of hidden states.

We propose here a hidden Markov LVM (HMLVM) with two major components including a conditional LVM and a continuation-ratio logit transition model. If we regard latent variables as random effects with certain structures, our model framework is similar to that proposed by Altman (2007). However, differences exist between our work and Altman's. First, the random effects in Alman (2007) mainly address the dependency of observations and are not of primary interest, whereas the latent variables in our model represent latent traits (e.g., psychological problems) that have specific meanings but cannot be characterized by a single observed variable. What's more, our conditional model reveals the effect of such latent traits on the outcome of interest. Second, given that cocaine-addiction conditions usually have a natural order from bad to good, our transition model for examining the bidirectional transition from one state to another is a continuation-ratio logit rather than a multinomial logit model. Third, Altman's computation method is not directly applicable to this study because our model involves additional latent quantities such as latent traits and missing data. Integrating them out leads to a prohibitively complex observed-data likelihood and infeasible computational burden.

We develop a Monte Carlo expectation-conditional-maximization (MCECM) procedure along with an efficient MCMC algorithm for parameter estimation. The asymptotic properties of parameter estimators are investigated. In addition, we take into account an important issue of testing the invariance of parameters across latent states, which is particularly relevant to the present study of cocaine use when we are interested in checking how the impacts of treatment and psychological problems on cocaine use vary across different cocaine-addiction conditions. However, to the best of our knowledge, no study has ever been conducted on the proposed model or on the associated theoretical developments.

The outline of this article is as follows. Section 2 introduces the HMLVM. Section 3 develops a MCECM procedure for estimation. The asymptotic properties of the parameter estimators and test statistics for checking the invariance of parameters are investigated. Section 4 presents simulation studies to examine the empirical performance of the proposed method. In Section 5, an application to the cocaine use data set is reported. Section 6 concludes the paper. Technical details are provided in Web Appendices.

2. Model Description

The model consists of two parts. In the first part, discussed in Section 2.1, the latent variable $\boldsymbol{\omega}_{it}$ allows for correlation within a given subject's response at a given time point. In the second part, discussed in Section 2.2, the latent state variable z_{it} allows for autocorrelation in a subject's responses over time. A graphical model presented in Figure 1 depicts the relations among the observed variables, latent factors, and the latent states associated with covariates and random effects under consideration. Here, the rectangles enclose observed variables and ellipses enclose latent factors and hidden states.

2.1 Conditional latent variable model

Consider the repeated measurements from N subjects across T occasions. Let y_{itj} denote the response of subject i at occasion t on the jth questionnaire item, $\boldsymbol{\omega}_{it}$ be a $p \times m$ vector of latent variables (factors), and z_{it} be categorical latent states taking values in a finite set $\mathbb{S} = \{1, \cdots, S\}$, where S is assumed known and fixed. The conditional LVM assumes a measurement model as follows:

$$[\mathbf{y}_{it}|z_{it}=s, \boldsymbol{\omega}_{it}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}_{\varepsilon}] \overset{\text{ind}}{\sim} N_p(\boldsymbol{\mu}^s+\boldsymbol{\Lambda}^s\boldsymbol{\omega}_{it}, \boldsymbol{\Psi}_{\varepsilon}^s), \qquad (1)$$

where $\mathbf{y}_{it} = (y_{it1}, \cdots, y_{itp})^\top$ is a $p \times 1$ vector of observed variables, $\boldsymbol{\mu}^s$ is a $p \times 1$ vector of intercepts, $\boldsymbol{\Lambda}^s$ is a $p \times m$ factor loading matrix, and $\boldsymbol{\Psi}_\varepsilon^s$ is a $p \times p$ diagonal matrix with diagonal elements $\psi_{\varepsilon k}^s$, $k = 1, \cdots, p$. The extension consisting in incorporating observed predictors into (1) is straightforward. The state-specific $\boldsymbol{\mu}^s$ and $\boldsymbol{\Psi}_\varepsilon^s$ allow for heterogeneity in grouping latent variables via observed variables over time. To examine the interrelationships among latent variables, we partition $\boldsymbol{\omega}_{it}$ into an $m_1 \times 1$ outcome latent vector $\boldsymbol{\eta}_{it}$ and an $m_2 \times 1$ explanatory latent vector $\boldsymbol{\xi}_{it}$ $(m_1 + m_2 = m)$. A structural equation is defined by

$$\boldsymbol{\eta}_{it}=\mathbf{B}^s\boldsymbol{\eta}_{it}+\boldsymbol{\Gamma}^s\boldsymbol{\xi}_{it}+\boldsymbol{\zeta}_{it}, \quad (2)$$

$$[\boldsymbol{\zeta}_{it}|z_{it}=s] \overset{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}_\zeta^s), \ [\boldsymbol{\xi}_{it}|z_{it}=s] \overset{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Phi}^s),$$

where $\mathbf{B}^s$ is a $m_1 \times m_1$ matrix of regression coefficients with the main diagonal elements being zero, $\boldsymbol{\Gamma}^s$ is a $m_1 \times m_2$ matrix of regression coefficients, $\boldsymbol{\Psi}_\zeta^s$ is a $m_1 \times m_1$ diagonal matrix with diagonal elements $\psi_{\zeta j}^s$, $j = 1, \cdots, m_1$, and $\boldsymbol{\Phi}^s$ is a $m_2 \times m_2$ covariance matrix. It is assumed that the processes of $\{\boldsymbol{\xi}_{it}\}$ and $\{\boldsymbol{\zeta}_{it}\}$ are independent.

In the conditional LVM defined by (1) and (2), the elements in $\boldsymbol{\eta}_{it}$ and $\boldsymbol{\xi}_{it}$ can be either latent factors or observed variables. When $m_1 = 1$, $\eta_{it} = y_{it1}$ implies that η_{it} is measured by (centralized) y_{it1} without error, and that appropriate constraints on $\boldsymbol{\mu}^s$, Λ^s, and $\boldsymbol{\Psi}_\varepsilon^s$ should be imposed. For instance, in the cocaine use study in Section 5, η_{it} (cocaine use) is measured by y_{it1} without error. Thus, $y_{it1}=\mu_1^s+\eta_{it}$, where μ_1^s is the mean of y_{it1} at state s, and the factor loading (λ_{11}^s) and the error variance $(\psi_{\varepsilon 1}^s)$ are set to 1 and 0, respectively.

Let $\mathbf{B}_0^s=\mathbf{I}_{m_1}-\mathbf{B}^s$, where $\mathbf{I}_{m_1}$ is the m_1-dimensional identity matrix. Based on the model assumptions and conditional on z_{it}, $\boldsymbol{\omega}_{it}$ has zero mean and covariance matrix

$$\boldsymbol{\Sigma}_\omega^s=\begin{bmatrix} (\mathbf{B}_0^s)^{-1}(\boldsymbol{\Gamma}^s\boldsymbol{\Phi}^s\boldsymbol{\Gamma}^{s^\top}+\boldsymbol{\Psi}_\zeta^s)(\mathbf{B}_0^s)^{-\top} & (\mathbf{B}_0^s)^{-1}\boldsymbol{\Gamma}^s\boldsymbol{\Phi}^s \\ \boldsymbol{\Phi}^s\boldsymbol{\Gamma}^{s^\top}(\mathbf{B}_0^s)^{-\top} & \boldsymbol{\Phi}^s \end{bmatrix}.$$

Through (1), the dependency among observed variables is explained by a substantially lower-dimensional latent vector $\boldsymbol{\omega}_{it}$. The correlation coefficient between y_{itk} and y_{itl}, given z_{it}, is

$$\mathbb{C}orr(y_{itk}, y_{itl}|z_{it}=s)=\frac{\boldsymbol{\Lambda}_k^{s^\top}\boldsymbol{\Sigma}_\omega^s\boldsymbol{\Lambda}_l^s}{\sqrt{\boldsymbol{\Lambda}_k^{s^\top}\boldsymbol{\Sigma}_\omega^s\boldsymbol{\Lambda}_k^s+\psi_{\varepsilon k}^s}\sqrt{\boldsymbol{\Lambda}_l^{s^\top}\boldsymbol{\Sigma}_\omega^s\boldsymbol{\Lambda}_l^s+\psi_{\varepsilon l}^s}}.$$

2.2 Continuation-ratio logit transition model

Let $\mathbf{z}_i = (z_{i1}, \cdots, z_{iT})^\top$ be the state sequence of subjects across the latent state space over time. A standard assumption in most HMMs assumes that $\{z_{it}\}$ follows the first order Markov chain with the transition probability given by

$$p(z_{it}=s|z_{i,t-1}=r)=q_{itrs},\ t=2,\cdots,T, \quad (3)$$

where q_{itrs} is the transition probability from state $z_{i,t-1}$ at occasion $t - 1$ to state z_{it} at occasion t for individual i.

Let $\mathbf{Q}_{it}$ be the $S \times S$ stochastic matrix with elements q_{itrs}. From (3), the joint distribution of $\mathbf{z}_i$ depends only on transition probabilities and the marginal distribution of the initial state. We assume that the initial distribution of z_{i1} is multinomial with $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_S)^\top$ such that $\nu_r \geqslant 0$ for $r = 1, \ldots, S$ and $\sum_{r=1}^{S}\nu_r=1.0$. Here, $\nu_1, \cdots, \nu_S$ can be treated as fixed if the panel length is large enough or estimated simultaneously with other model parameters. Alternative of initial distribution can be chosen as the limit (stationary) distribution of z_{it} provided that z_{it} is stationary or taken as a point mass $\delta_{z0}(\cdot)$ for some preassigned value z_0. This paper allows for heterogeneity of the transition probability of the hidden Markov chain by incorporating subject- and/or occasion-specific fixed and random effects.

Assuming that the states $\{1, \cdots, S\}$ in $\mathbb{S}$ are ordered, the transition probabilities can then be modeled through the following continuation-ratio logit model (Agresti, 2002; Ip et al., 2013). Specifically, for $t = 2, \cdots, T$ and $s = 1, \cdots, S - 1$, we have

$$\log\left(\frac{p(z_{it}=s|z_{i,t-1}=r)}{p(z_{it}>s|z_{i,t-1}=r)}\right)=\log\left(\frac{q_{itrs}}{q_{itrs,s+1}+\cdots+q_{itrS}}\right)=\eta^*_{itrs}=\alpha_{rs}+\mathbf{w}_{it}^\top\boldsymbol{\beta}+\mathbf{v}_{it}^\top\mathbf{b}_i, \quad (4)$$

where α_{rs} is a state-specific intercept, $\mathbf{w}_{it}$ and $\mathbf{v}_{it}$ are, respectively, $\kappa_1 \times 1$ and $\kappa_2 \times 1$ vectors of covariates for individual i at occasion t, $\boldsymbol{\beta}$ is a $\kappa_1 \times 1$ vector of common fixed effects coefficients, and $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$ is a $\kappa_2 \times 1$ vector of subject-specific random effects. The parameterization in (4) is intended to facilitate interpretation of transition to a state rather than a better one. Let $\varpi_{itrs} = p(z_{it} = s|z_{it} \geqslant s, z_{i,t-1} = r)$, the continuation-ratio logits in the left-hand side of (4) can be written as $\log[\varpi_{itrs}/(1 - \varpi_{itrs})] = \text{logit}(\varpi_{itrs})$ (Agresti, 2002, p. 289). Thus, $\boldsymbol{\beta}$ and other parameters in (4) can be interpreted similarly as those in the logit model. By introducing random effects into (4), the hidden process is no longer Markovian.

2.3 Model identifiability

There are two model indeterminacies in the proposed HMLVM. One is from the invariance of the covariance matrix of latent factors under orthogonal transformation in the measurement equation (1). We follow the common practice in LVM literature to fix appropriate elements of the factor loading matrix at preassigned values to solve this problem. The other is related to label switching, which causes a difficulty in parameter estimation because the resulting likelihood will be multi-modal. We use the method proposed by Scott et al. (2005) to implement our algorithm without constraint but use cross-validation methods to explore the initial values of estimates (see Section 5).

3. Statistical Inference

3.1 ML estimation via MCECM

For $i = 1, \cdots, N$, let $\mathbf{y}_i=(\mathbf{y}_{i1}^\top,\cdots,\mathbf{y}_{iT}^\top)^\top$ be a $(Tp) \times 1$ vector of observations across T occasions for subjects i and $\boldsymbol{\omega}_i=(\boldsymbol{\omega}_{i1}^\top,\cdots,\boldsymbol{\omega}_{iT}^\top)^\top$ be a $(Tm) \times 1$ vector of latent factors that are associated with $\mathbf{y}_i$. The observed-data log-likelihood function is

$$l_i(\boldsymbol{\theta})=\log\left[\int_{\mathbb{S}^T\times\mathbb{R}^{\kappa_2}}\left\{\int_{\mathbb{R}^m}p(\mathbf{y}_i|\boldsymbol{\omega}_i,\mathbf{z}_i,\boldsymbol{\theta})p(\boldsymbol{\omega}_i|\mathbf{z}_i,\boldsymbol{\theta})d\boldsymbol{\omega}_i\right\}p(\mathbf{z}_i|\mathbf{b}_i,\boldsymbol{\theta})p(\mathbf{b}_i|\boldsymbol{\Sigma}_b)\mu^T(d\mathbf{z}_i)d\mathbf{b}_i\right], \tag{5}$$

where $\boldsymbol{\theta}$ is the vector of all unknown parameters, and $\mu^T(A) = \sum_{\mathbf{z}\in A}\delta_{\mathbf{z}}$ is the counting measure on the product space $\mathbb{S}^T = \mathbb{S} \times \cdots \times \mathbb{S}$.

Latent factors $\boldsymbol{\omega}_i$ involved in $l_i(\boldsymbol{\theta})$ can be integrated out in the context of linear LVM. Nevertheless, direct maximization of $l(\boldsymbol{\theta})$ is still computationally infeasible because it also involves a complicated high-dimensional integration with respect to latent states $\mathbf{z}_i$ and random effects $\mathbf{b}_i$. Altman (2007) suggested a MCEM algorithm, of which the E-step was implemented by drawing observations from the prior distribution of random effects and the M-step was carried out via numerical maximization. However, Altman's method is not directly applicable here because our study involves additional latent quantities such as latent factors and missing questionnaire data.

We propose the use of MCECM algorithm to obtain the estimation of model parameters. The E-step is implemented by drawing observations from the joint posterior distribution of the latent quantities, and the M-step combines the conditional maximization and Newton-Raphson algorithm. Let $\mathbf{Y} = \{\mathbf{y}_1, \cdots, \mathbf{y}_N\}$, $\mathbf{Z} = \{\mathbf{z}_1, \cdots, \mathbf{z}_N\}$, $\boldsymbol{\Omega} = \{\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_N\}$, and $\mathbf{B} = \{\mathbf{b}_1, \cdots, \mathbf{b}_N\}$. We treat $\{\mathbf{Z}, \boldsymbol{\Omega}, \mathbf{B}\}$ as hypothetical missing data and augment them with $\mathbf{Y}$ to approximate the conditional expectation in the E-step. The Gibbs sampler is implemented to sample from $p(\mathbf{Z}, \boldsymbol{\Omega}, \mathbf{B}|\mathbf{Y}, \boldsymbol{\theta})$ iteratively through (a) generating $\mathbf{Z}$ from $p(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\Omega}, \mathbf{B}, \boldsymbol{\theta})$, (b) generating $\boldsymbol{\Omega}$ from $p(\boldsymbol{\Omega}|\mathbf{Y}, \mathbf{Z}, \mathbf{B}, \boldsymbol{\theta})$, and (c) generating $\mathbf{B}$ from $p(\mathbf{B}|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\Omega}, \boldsymbol{\theta})$. For notational simplicity, we do not incorporate missing questionnaire data here. The proposed procedure can be extended to accommodate missing questionnaire data without difficulty. In the extension, the likelihood function involves the observed data and the missing indicators, resulting in a joint likelihood. The details of the MCECM algorithm and the extension consisting in incorporating missing questionnaire data are described in Web Appendix A.

An important issue regarding the convergence of EM algorithm is that it may converge to a local maxima or even to a saddle point. We thus adopt the method proposed by McLachlan and Peel (2000) to choose the maximum likelihood (ML) estimate that results in the highest log-likelihood among all non-spurious solutions obtained via using the MCECM algorithm with different starting values of parameters. In the present study, three groups of different starting values are used in the numerical studies. Let $\hat{\theta}_N$ be the ML estimate of θ obtained using the MCECM algorithm. Theorem 1 in Web Appendix B investigates the asymptotic properties of $\hat{\theta}_N$. The asymptotic covariance matrix of $\hat{\theta}_N$ is given by the inverse of

$$N\boldsymbol{J}(\hat{\boldsymbol{\theta}}_N) \approx \sum_{i=1}^{N}\left[\frac{\partial l_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial l_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_N}.$$

3.2 Model selection and hypothesis test

In the proposed HMLVM, determining the number of latent states, S, is fundamental and should be taken into consideration. Information criteria, such as Akaike information criteria (AIC) and Bayesian information criteria (BIC), have been widely used for model selection, especially in the context of non-nested models. In this study, we use AIC and BIC to compare HMLVMs with different numbers of latent states. The computation of AIC and BIC is provided in Web Appendix A.

After the number of latent states is determined, test of heterogeneity in model parameters across different latent states is likewise of interest. For LVM, a common test is made on the invariance of the factor loadings. Consider the following hypotheses:

$$H_0{:}\boldsymbol{\Lambda}^1{=}\cdots{=}\boldsymbol{\Lambda}^S \text{ v.s. } H_1{:} \text{ there exists at least a } (r,s) \text{ such that } \boldsymbol{\Lambda}^r \neq \boldsymbol{\Lambda}^s. \quad (6)$$

We propose the Wald and Score test statistics to perform the hypothesis testing. Theorem 2 in Web Appendix B presents their asymptotical Chi-square distributions. This result can be easily extended to test the invariance of others parameters, such as $\boldsymbol{\mu}^s$ and $\boldsymbol{\Gamma}^s$ in LVM, or to test whether some elements of $\boldsymbol{\gamma}$ are equal to zero in the transition model. Notably, for the variance parameter $\boldsymbol{\sigma}$ ($\geqslant 0$) [see Equation (10) below], the test of $H_0 : \boldsymbol{\sigma} = 0$ is a boundary problem, and thus the standard Wald test is not valid in this case. One can instead conduct a model comparison between a random effect model ($\boldsymbol{\sigma} > 0$) and a fixed effect model ($\boldsymbol{\sigma} = 0$).

4. Simulation Study

In this section, we conduct a simulation study to assess the empirical performance of the proposed methodology described in Section 3.

4.1 Simulation 1

We first investigate the finite sample performance of the estimation procedure under different choices of N and T. Let $\mathbf{y}_{it} = (y_{it1}, \cdots, y_{it9})^\top$, $\boldsymbol{\omega}_{it} = (\eta_{it}, \xi_{it1}, \xi_{it2})^\top$, and $\mathbb{S} = \{1, 2,$

3}. The measurement equation is defined by (1) with $p = 9$, $m = 3$, $m_1 = 1$, $m_2 = 2$, $\boldsymbol{\mu}^s=(\mu_1^s,\cdots,\mu_9^s)^\top$, $\boldsymbol{\Psi}_\varepsilon^s=\text{diag}\{\psi_{\varepsilon 1}^s,\cdots,\psi_{\varepsilon 9}^s\}$,

$$\boldsymbol{\Lambda}^s=\begin{bmatrix} 1 & \lambda_{21}^s & \lambda_{31}^s & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda_{52}^s & \lambda_{62}^s & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{83}^s & \lambda_{93}^s \end{bmatrix}^\top, \quad \text{and } \boldsymbol{\Phi}^s=\begin{bmatrix} \phi_{11}^s & \phi_{12}^s \\ \phi_{21}^s & \phi_{22}^s \end{bmatrix},$$

where the ones and the zeros are fixed to identify the model. The structural equation is given by $\eta_{it}=\Gamma_1^s\xi_{it1}+\Gamma_2^s\xi_{it2}+\zeta_{it}$ with $[\zeta_{it}|z_{it}=s]\sim N(0,\psi_\zeta^s)$. The true population values of the unknown parameters are presented in Table 1.

For the transition model, we assume that the initial distribution is $\nu = (1, 0, 0)^\top$. Consider the continuation-ratio logit model

$$\eta_{itrs}^*=\alpha_{rs}+\mathbf{w}_{it}^\top\boldsymbol{\beta}+\upsilon_{it}b_i,\ r=1,2,3,\ s=1,2, \quad (7)$$

where $\mathbf{w}_{it}^\top=(w_{it1},w_{it2})$ is a vector of fixed covariates, w_{it1} and w_{it2} are, independently, drawn from Bernoulli distribution with success probability 0.3, and b_i is generated from $N(0, \sigma^2)$ with $\sigma^2 = 1.0$, and $\upsilon_{it} = 1.0$. The covariates w_{it1} and w_{it2} are then held constant over the subsequent MCECM iterations. The true values of the parameters are presented in Table 1.

The MCECM algorithm is implemented for parameter estimation. We chose the starting values through disturbing the true values of the unknown parameters in θ by adding 1.0 to the intercept and regression parameters and by multiplying 1.5 to the variance parameters. In the use of MCMC methods for approximating conditional expectations, we collected 200 observations after deleting 200 samples as the burn-in at the first 10 EM iterations and then increased the sample size by 50 times for the latter iterations. The adaptive Metropolis rejection algorithm was employed for the sampling of b_i. At each iteration, the one-step Newton-Raphson algorithm was implemented for updating the parameters in the transition model. The convergence of the MCECM algorithm was monitored via a plot of the observed-data log-likelihood function against the number of iterations.

To evaluate the finite sample performance of the ML estimates, we considered four scenarios with $(N, T) = (300, 4)$, $(1000, 4)$, $(300, 10)$, and $(1000, 10)$, respectively. For each scenario, the simulation is conducted on the basis of 100 replications. We first conducted a few test runs to get a rough idea about the number of iterations at convergence. The pilot study showed that the MCECM algorithm converged within 30 iterations for all scenarios. To be conservative, we took 35 iterations in each replication. The result obtained under $(N, T) = (300, 4)$ is reported in Table 1. The values of root mean square error (RMS) and average approximate standard error (SE) are close to zero. Most of the coverage rates of the estimators are slightly higher than the nominal level (95%), implying that the standard errors of the estimators tend to be slightly overestimated in this case. The performance of parameter estimates, their standard error estimates, and the coverage rates is improved as the sample size and/or the panel length increase. The details are reported in Web Appendix C.

4.2 Simulation 2

To assess the finite sample performance of the information criteria in determining the number of latent states, we generated four data sets based on the same LVM as defined in Simulation 1 but the transition model with 1-state, 2-state, 3-state, and 4-state, respectively. The four scenarios in Simulation 1 are again considered. The transition model is defined by (7) except that when $S = 2$, $\boldsymbol{\nu} = (0.2, 0.8)^{\top}$, $\alpha_{1s} = 0.3$, and $\alpha_{2s} = 0.5$ for $s = 1$; when $S = 3$, $\boldsymbol{\nu} = (0.2, 0.1, 0.7)^{\top}$, $\alpha_{1s} = 0.3$, $\alpha_{2s} = 0.5$, and $\alpha_{3s} = 0.7$ for $s = 1, 2$; and when $S = 4$, $\boldsymbol{\nu} = (0.2, 0.1, 0.3, 0.4)^{\top}$, $\alpha_{1s} = 0.3$, $\alpha_{2s} = 0.5$, $\alpha_{3s} = 0.7$, and $\alpha_{4s} = 0.9$ for $s = 1, 2, 3$. The MCECM algorithm is implemented to obtain the ML estimates of the model parameters under each setting. To compute the observed-data log-likelihood function, we use the Gaussian-quadrature numerical method with 100 knots to approximate the integrals involved. Table 2 reports the results based on 100 replications, indicating that AIC and BIC generally perform satisfactorily and their performance improves as N or T increases. To examine whether other factors such as the Monte Carlo sample size in the E-step and the number of knots in the Gaussian-quadrature approximation of the M-step might affect the distribution of the estimators as well as the values of AIC and BIC, we disturb the Monte Carlo sample size from 200 to 1000 and the number of knots from 100 to 50. The estimation and model selection results are similar and not reported.

In summary, a size of $(N, T) = (300, 4)$ provides reasonable estimation and model selection results. The increase of N and/or T would reduce the values of RMS and SE, improve the coverage rates of the estimators, and enhance the performance of AIC and BIC. This study uses 200 Monte Carlo samples and 100 knots in Gaussian quadrature method in each of the scenarios considered. The increase of these factors does not significantly improve the finite sample performance of the MCECM algorithm.

The computing time for obtaining the results of Tables 1 and 2 in each replicate takes 30 minutes using visual C++ for window 7 with cpu clock speed at 2.93Hz. The computer code is available online. We conduct Simulation 3 to examine the empirical performance of the test statistics proposed in Theorem 2. The details are provided in Web Appendix C.

5. A Longitudinal Study of Cocaine Use

In this section, we use the proposed method to analyze the cocaine use data set described in the Introduction. The data set was collected from 321 patients at baseline, one year, two years, and 12 years after treatment ($t = 0, 1, 2, 3$), in which some patients were confirmed to be deceased (8.7%), some declined to be interviewed, and some were either out of the country or too ill to be interviewed. Consequently, there is a large amount of missing questionnaire data in this longitudinal data set. The questionnaire items include y_1 =Days of cocaine use per month (CC), y_2 = Times per month in formal treatment (outxfreq), y_3 = Months in formal treatment (outTXmon), y_4 = beck inventory (BI), y_5 = depression (DEP), and y_6 = anxiety (AN). Among them, y_1 reects the participants' cocaine use severity, which measures the outcome variable η without error; $\{y_2, y_3\}$ are all related to treatment received by participants, and are therefore grouped into a latent trait "treatment (ξ_1)"; and $\{y_4, y_5, y_6\}$ all characterize mental health-related condition, and are therefore grouped into another latent trait "psychological problems (ξ_2)". In these variables, $y_1, \cdots, y_6$ are observed, ξ_1 and

ξ_2 are latent, and $\eta = y_1$. Based on the questionnaire, $\{y_4, y_5, y_6\}$ are continuous, and $\{y_1, y_2, y_3\}$ take integer values in the ranges of [0, 30], [0, 30], and [0, 12], respectively. Given their relatively large ranges, $\{y_1, y_2, y_3\}$ are regarded as continuous as well. The main goal of this study is to investigate the effects of treatment (ξ_1) and psychological problems (ξ_2) on cocaine use (η) and simultaneously examine the change patterns of these effects during different episodic cocaine-addiction periods.

We considered the measurement equation defined in (1) with the factor loading matrix

$$\mathbf{\Lambda}^s = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda_{32}^s & 0 & 0^* & 0 \\ 0 & 0 & 0 & 1 & \lambda_{53}^s & \lambda_{63}^s \end{bmatrix}^\top, \quad (8)$$

where the ones and the zeros were treated as fixed parameters. Given that η corresponds to a single observed variable ‘CC’, we fixed λ_{11}^s at 1.0 and the corresponding unique error $\psi_{\varepsilon 1}^s$ at zero. To examine the effects of treatment and psychological problems on cocaine use, we considered a linear structural equation

$$\eta = \Gamma_1^s \xi_1 + \Gamma_2^s \xi_2 + \zeta. \quad (9)$$

Given the reversible feature of cocaine use, it is pertinent to conceptualize it as a stochastic process transitioning between different cocaine-addiction states ($z_{it} = 1, \cdots, S$) that have a natural order from bad to good. Unlike the continuous cocaine use variable η in (9), z_{it} is a discrete latent variable, representing hidden states that cocaine addicts may go through over time. We adopted mixed continuation-ratio logit model (4) to examine the bidirectional transition between various cocaine-addiction states over time. A baseline measurement “currently employed at intake (w_i)” (1 – unemployment, 0 – otherwise) was included as a covariate. To take into account the effect of unequally spaced time intervals on transition probabilities, we incorporated the time intervals into (4) as follows:

$$\eta_{itrs}^* = \alpha_{rs} + w_i \beta + \upsilon_{it} b_i, \quad (10)$$

where $b_i \sim N(0, \sigma^2)$, and $\upsilon_{it} = 0, 1, 2$, and 12. Here, the random effect b_i explains the dependence of repeated measurements for patient i at different occasions, and the time-varying covariate υ_{it} raises such dependence over time. The rationale behind this assumption is that the transition pattern would be increasingly dependent on patient-specific characteristics rather than on the baseline employment status over time.

Because of the existence of missing data, we first determine an appropriate missing data mechanism between missing at random (MAR) and missing not at random (MNAR). In the presence of missing data, we use the idea of data augmentation by augmenting the observed data with latent quantities including latent state variables, random effects, latent factors, and missing data. Logistic regressions

$\text{logit}(p(r_{itj}=1|\mathbf{y}_i,\boldsymbol{\varphi})=\varphi_0+\sum_{t=1}^{T}\sum_{j=1}^{p}\varphi_{tj}y_{itj}=\varphi_0+\boldsymbol{\varphi}^\top\mathbf{y}_i$ and $\text{logit}(p(r_{itj}=1|\mathbf{y}_i,\boldsymbol{\varphi})=\varphi_0+\sum_{t=1}^{T}\sum_{j=1}^{p}\varphi_{tj}(1-r_{itj})y_{itj}=\varphi_0+\boldsymbol{\varphi}^\top\mathbf{y}_{i,\text{obs}}$ are used to model the missing probability of y_{itj} under MNAR or MAR, respectively, where r_{itj} is a missing indicator variable taking value 1 if y_{itj} is missing and 0 otherwise, $\boldsymbol{\varphi} = (\varphi_{11}, \cdots, \varphi_{1p}, \cdots, \varphi_{T1}, \cdots, \varphi_{Tp})^\top$, and $\mathbf{y}_{i,obs}$ includes the observed elements of $\mathbf{y}_i$. Although the two missing data models result in distinct conditional distributions of missing data, the associated observed-data likelihood functions both involve observations in $\mathbf{y}_i$ and the corresponding missing indicators. Thus, the likelihood-based criteria can be used to compare the proposed MNAR and MAR mechanisms. Notably, the log-likelihoods in the 1-state and s-state ($s > 1$) models are not computed on exactly the same observed data because the latter involves data augmentation using additional covariates w_i and υ_{it}. Thus, instead of using AIC and BIC, we examined the existence of heterogeneity by plotting the histograms and estimated predictive distributions of the observed variables. Figure 2 presents the histograms of y_1, y_2, and y_3, and their predictive densities estimated in the 1-state and 2-state models. Apparently, the predictive densities estimated in the 2-state model captured the patterns of the histograms but those estimated in the 1-state model did not. Thus, the 1-state model is inadequate and should not be considered in this study. Then, we used AIC and BIC to compare 2-state, 3-state, and 4-state models with MAR or MNAR assumption. When fitting the data set with a 4-state model, we found many heywood cases (i.e., $\psi_{\varepsilon j}^{s}\leqslant 0$ for some j and s). This phenomenon usually occurs when data contain outliers or sample size is not large enough (Lee and Xia, 2006). In the present study, a possible reason for heywood cases occurrence is that the insufficient samples make certain states lack observations, thereby leading to the MCECM algorithm unstable or divergent. To avoid heywood cases, we fixed $\psi_{\varepsilon j}^{s}=1$ in the 4-state model. We performed the comparison on 2-state, 3-state, and 4-state models with or without random effect and under MAR or MNAR assumption. The results are summarized in Table 3. The 3-state model with random effect and MAR missing is among the best. We then fixed the number of hidden states at 3 and regarded the missing data as MAR. Based on the common knowledge about cocaine addiction process, we interpreted the 3 states as addiction to, moderate dependence on, and swearing off cocaine. We then focused the subsequent inference on (i) obtaining parameter estimates at each of the 3 states, (ii) examining the transition probabilities among the 3 states, and (iii) testing the invariance of the factor loadings and regression coefficients across the 3 states.

To obtain good starting values of parameters, we employed the permutation sampler (Frühwirth-Schnatter, 2001) to conduct a Bayesian analysis, and then took the Bayesian estimates as the starting values. Table 4 reports the parameter estimates, their standard error estimates (in parentheses), and the corresponding P-values in the significance test. For the sake of comparison, we standardized the distributions of ξ_1 and ξ_2 so that their variances equal 1 at each state, and then transformed the regression coefficients and other parameters accordingly. The three cocaine-addiction states and the state-specific effects of treatment and psychological problems on cocaine use are interpreted on the basis of transformed estimates as follows. State 1 represents a severe addiction state, wherein patients are dependent on cocaine both physically and mentally. The result of $\hat{\Gamma}_1^1=-0.316\ (0.101)$ and

$\hat{\Gamma}_2^1=0.070\;(0.028)$ implies that treatment and psychological problems all influence cocaine use and the treatment effect seems more pronounced. More treatments and less psychological problems (or better mental health) would be substantially beneficial to the control of cocaine use. State 2 represents a moderately cocaine-dependent state, wherein patients depend on cocaine less physically but more mentally compared to those in state 1. The result of $\hat{\Gamma}_1^2=-0.135\;(0.072)$ and $\hat{\Gamma}_2^2=0.145\;(0.031)$ indicates that the effect on cocaine use becomes less significant for treatment but more significant for individuals' psychological problems than that in state 1. State 3 indicates a minor addiction state, wherein patients suffer the least from cocaine addiction. The result of $\hat{\Gamma}_1^3=-0.074\;(0.035)$ and $\hat{\Gamma}_2^3=0.354\;(0.103)$ shows that the effect of psychological problems on cocaine use becomes even stronger compared to those in states 1 and 2. Family support, friendship, and environment might be more important than formal treatment for cocaine-dependents in this state. In the mixed-effect transition model, $\hat{\beta}$ = 0.374 (0.185) can be interpreted as follows: the estimated odds of transitioning from a state r at time $t-1$ to a state s at time t rather than to a better state ($z_{it} > s$) at time t for addicts unemployed at intake are exp(0.374) = 1.454 times the estimated odds for addicts employed at intake. Thus, having a job can increase the probability of cocaine users transitioning from a state to a better one. The highly significant variance estimate $\hat{\sigma}^2$ = 0.891 (0.128) reveals great heterogeneity (or high dependence) in transitions from one state to another for the same subject at different occasions. We also conducted an analysis using a fixed-effect transition model. The result is different and not reported. In particular, $\hat{\Gamma}_1^1=0.174\;(0.016)$ and $\hat{\Gamma}_1^2=0.058\;(0.017)$ indicate that for severe or moderate cocaine-addicts, medical treatment would increase their cocaine use, whereas $\hat{\Gamma}_2^2=-0.073\;(0.019)$ implies that more psychological problems (or worse mental health) would lessen cocaine-addiction. These confusing results may reveal the danger of ignoring possible heterogeneity or dependency in modeling the transition process.

The estimated factor loadings can be interpreted as follows. In state 1, $\lambda_{31}^1=0.404\;(\text{fixed})$ and $\hat{\lambda}_{32}^1=0.174\;(0.029)$ imply that outxfreq (y_2) and outTXmon (y_3) significantly contribute to the characterization of treatment (ξ_1) in the same direction but the contribution is relatively smaller for y_3 than for y_2. Similarly, $\lambda_{43}^1=0.527\;(\text{fixed})$, $\hat{\lambda}_{53}^1=0.970\;(0.031)$, and $\hat{\lambda}_{63}^1=0.844\;(0.042)$ imply that BI (y_4), DEP (y_5), and AN (y_6) all significantly contribute to the characterization of psychological problems (ξ_2) in the same direction but the contribution is relatively larger for y_5 and y_6 than for y_4. Further, $\hat{\lambda}_{32}^s$, $\hat{\lambda}_{53}^s$, and $\hat{\lambda}_{63}^s$, respectively, decrease to 0.199 (0.073), 0.724 (0.089), and 0.681 (0.078) in State 2, as well as 0.036 (0.001), 0.637 (0.069), and 0.504 (0.068) in State 3, indicating that the associations between y_3 and ξ_1 as well as $\{y_5, y_6\}$ and ξ_2 decrease as the state transits from bad to good.

Figure 3 depicts the optimal state sequence for each individual. Let $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{mis}$ be the sets of the observed and missing questionnaire data, respectively. The optimal path of transition for subject i is defined as $\mathbf{z}_{\text{iopt}}=\arg\max_{z_{i1},\cdots,z_{iT}} E[p(z_{i1},\cdots,z_{iT}|\boldsymbol{Y},\boldsymbol{\Omega},\mathbf{B},\hat{\boldsymbol{\theta}})|\boldsymbol{Y}_{\text{obs}},\hat{\boldsymbol{\theta}})]$, in which the maximization is taken in the state space with 3^4 = 81 points for each subject, and the expectation is taken with respect to $p(\mathbf{Y}_{mis}, \mathbf{\Omega}, \mathbf{B}|\mathbf{Y}_{obs}, \hat{\theta})$. We used Monte Carlo method to compute the involved probabilities via drawing 10,000 observations of $\{\mathbf{Y}_{mis}, \mathbf{\Omega}, \mathbf{B}\}$ from $p(\mathbf{Y}_{mis}, \mathbf{\Omega}, \mathbf{B}|\mathbf{Y}_{obs}, \hat{\theta})$ with 10,000 burn-ins deleted. The frequencies of State 1, 2, and 3 at time 1, 2, 3, and 4 are {0.91, 0.0, 0.19}, {0.36, 0.14, 0.50}, {0.45, 0.04, 0.51}, and {0.26, 0.02, 0.72}, respectively. This implies the following transition tendency of the underlying states. At baseline, a majority (91%) and minority (19%) of patients are in states 1 and 3, respectively. After one year's treatment, apparent transitions from state 1 to states 2 and 3 result in 45% reduction but 14% and 31% addition of patients in States 1, 2, and 3. This tendency is not as that significant thereafter. In one year to two years, most patients' states keep unchanged except for a slight rebound from State 2 to 1. After 12 years treatment, the proportion of patients decreases to 26% for State 1 and increases to 72% for State 3.

Finally, we tested the invariance of the factor loadings and regression coefficients in LVM using the proposed Wald test statistics(T_N^W) and Score test statistics(T_N^{SC}). The null hypothesis is specified as $H_0 : \mathbf{\Lambda}^1 = \mathbf{\Lambda}^2 = \mathbf{\Lambda}^3, \mathbf{\Pi}^1 = \mathbf{\Pi}^2 = \mathbf{\Pi}^3$. The values of T_N^W and T_N^{SC} are equal to 132.43 and 32.67, respectively, demonstrating a strong evidence (at 0.05 level of significance) of heterogeneity in factor load matrix $\mathbf{\Lambda}^s$ and regression coefficient matrix $\mathbf{\Pi}^s$.

6. Discussion

In this paper, a HMLVM has been proposed to analyze multivariate longitudinal data. We have developed a ML procedure, coupled with the MCECM algorithm, to carry out statistical inference. We have proposed hypothesis testing approach to test the invariance among parameters across different states. Although the existing softwares, such as Mplus (Muthén, 2013), can be used to analyze dynamic LVMs, they are not directly applicable to this study because of the inclusion of additional latent quantities in the proposed model.

The present work has limitations. First, the proposed model assumes that the serial correlation in y is modeled by the latent state z only. This assumption may be restrictive in practice and could be released by incorporating other components into the conditional model. Second, the amount of parameters involved is in general huge compared to the amount of data and may seriously limit the model assets on these data. Thus, the proposed method should be used with caution in the case of small sample size. Third, the convergence of MCECM algorithm was monitored via the plot of log-likelihoods against the number of iterations. Compared to the complexity of the proposed model and possibly flat areas of the log-likelihood, this criterion may be weak. An alternative approach is to monitor convergence by computing the relative error of parameter estimates (Lee and Song, 2004). Fourth, in the real application of Section 5, the time intervals between different occasions are unequally spaced: the time interval between $t = 2$ and $t = 3$ is 10, whereas those between other adjacent occasions are 1. We considered the use of 10-step transition probabilities to describe transitioning between $t = 2$ and $t = 3$, but the computer program broke down due to lack of sufficient data. Thus, great caution should be exercised in the presence of completely unbalanced data because models too complex may easily become unidentifiable or intractable. Finally, the proposed model assumes linear LVM and continuous responses, an extension to accommodate nonlinear LVM and discrete data is of great interest. These extensions raise theoretical and computational challenges and further investigation is needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was partially supported by GRF grants 14305014 and 14601115 from the Research Grant Council of HKSAR, direct grants from CUHK, NSFC grant 11471277 from China, NIH grants MH086633 and 1UL1TR001111, and NSF grants SES-1357666 and DMS-1407655. The authors thank the Editor, the Associate Editor, and two reviewers for their valuable comments which improved the paper substantially, and Dr. Yih-Ing Hser for providing the data set in the real example.

References

Altman RM. Mixed hidden Markov models. Journal of the American Statistical Association. 2007; 477:201–210.

Agresti, A. Categorical Data Analysis. New York: John Wiley & Sons; 2002.

Bartolucci F, Farcomeni A. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. Journal of the American Statistical Association. 2009; 104:816–831.

Bartolucci F, Lupparelli M, Montanari GE. Latent Markov modelfor longitudinal binary data: an application to the performance evaluation of nursing homes. The Annals of Applied Statistics. 2009; 3:611–636.

Bartolucci, F., Farcomeni, A., Pennoni, F. Latent Markov models for longitudinal data. Chapman & Hall/CRC, Taylor and Francis Group; 2013.

Cappé, O., Moulines, E., Rydén, T. Inference in Hidden Markov Models. New York: Springer; 2005.

Chow SM, Grimm KJ, Filteau G, Dolan CV, McArdle J. Regime-switching bivariate dual change score model. Multivariate Behavioral Research. 2013; 48:463–502. [PubMed: 26742002]

Frühwirth-Schnatter S. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. Journal of the American Statistical Association. 2001; 96:194–209.

Ip EH, Zhang Q, Rejeski WJ, Harris TB, Kritchevsky S. Partially ordered mixed hidden Markov model for the disablement process of older adults. Journal of the American Statistical Association. 2013; 108:370–384. [PubMed: 24058222]

Lee SY, Song XY. Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. Biometrics. 2004; 60:624–636. [PubMed: 15339284]

Lee SY, Xia YM. Maximum likelihood methods in treating outliers and symmetrically heavy-tailed distributions for nonlinear structural equation models with missing data. Psychometrika. 2006; 71:565–585.

Maruotti A. Mixed hidden markov models for longitudinal data: An overview. International Statistical Review. 2011; 79:427–454.

McLachlan, GJ., Peel, D. Finite mixture models. New York: John Wiley & Sons; 2000.

Muthén B. Mplus Version 7 User's Guide. 2013

Scott SL, James GM, Sugar CA. Hidden Markov models for longitudinal comparisons. Journal of the American Statistical Association. 2005; 100:369–369.

Vermunt JK, Rolf L, Ulf B. Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. Journal of Educational and Behavioral Statistics. 1999; 24:179–207.

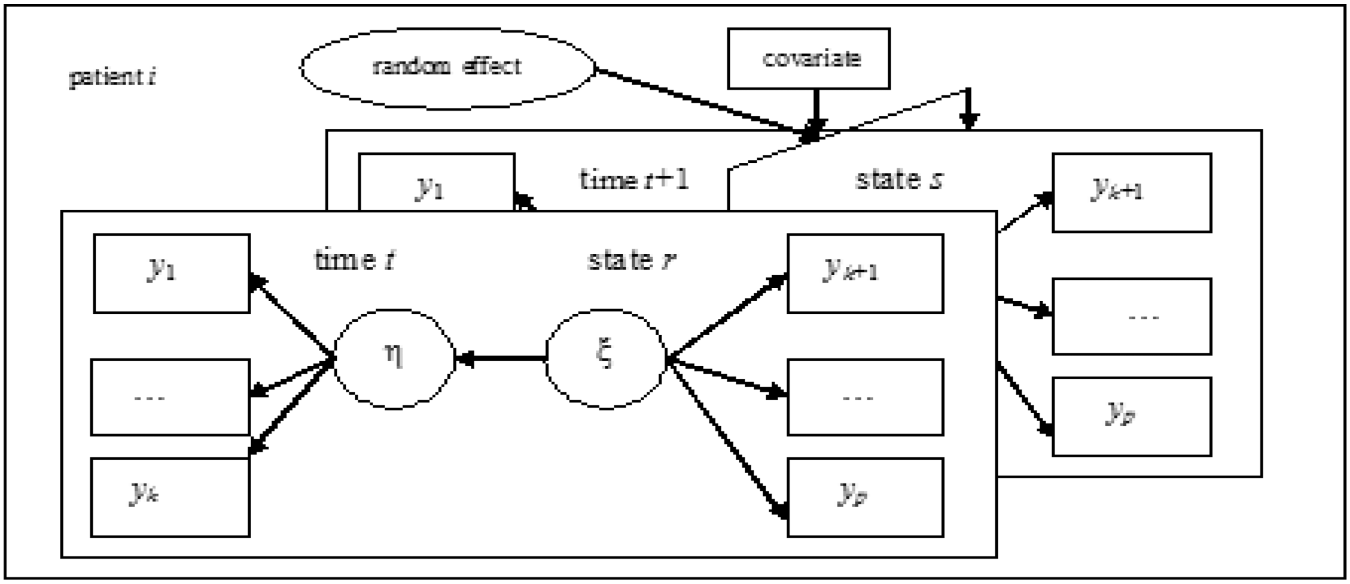

Figure 1.

Path diagram of the proposed HMLVM: The rectangles represent the observed responses or fixed covariates, and the ellipses denote the unobserved latent factors or random effects. The arrows identify the direct effect or transition between two random quantities.

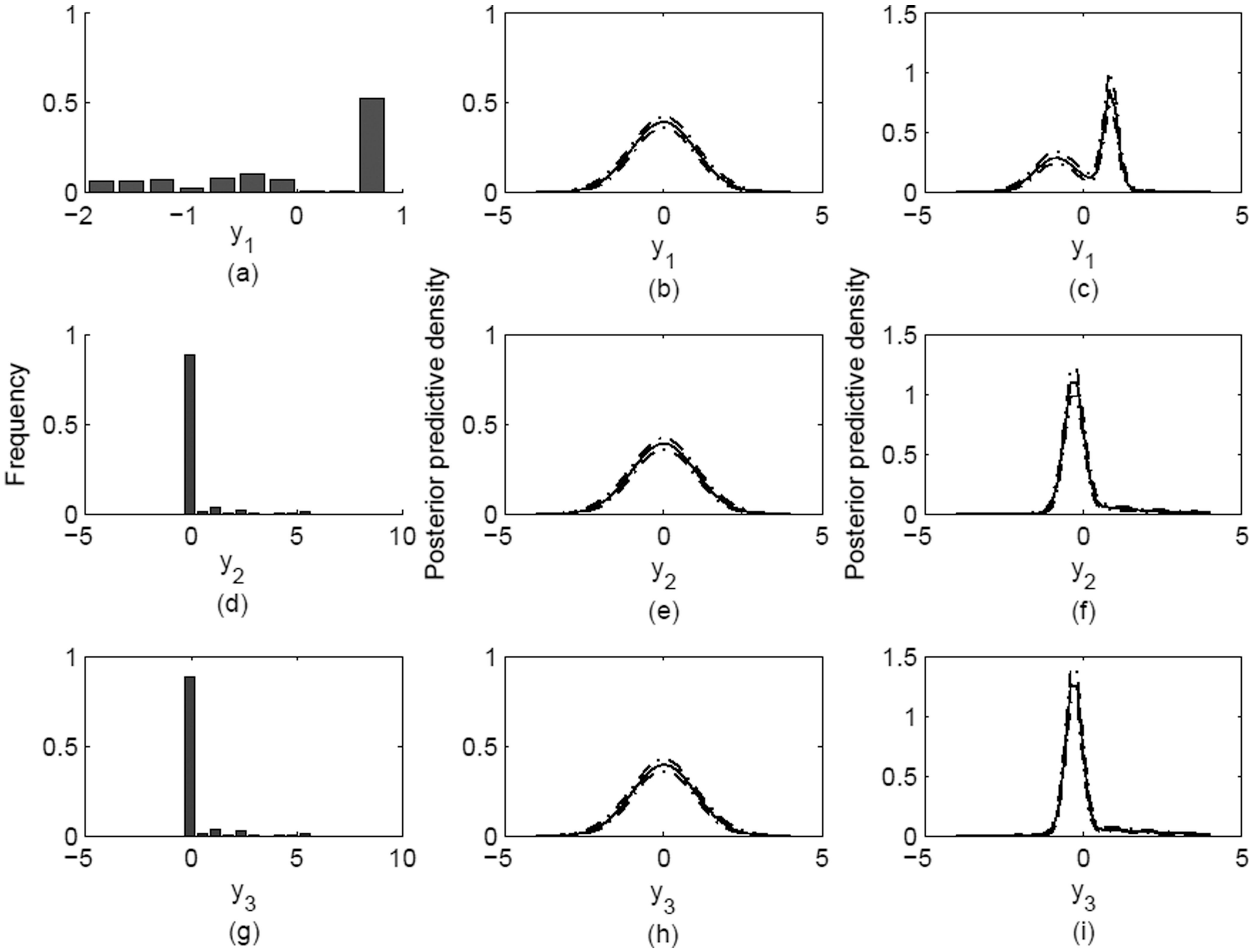

Figure 2.

The histograms and estimated predictive densities of CC (y_1), Outxfreq (y_2), and OutTXmon (y_3) at baseline. From top to bottom, Column 1: the histograms of y_1, y_2, and y_3; Column 2: the estimated predictive densities of y_1, y_2, and y_3 in a 1-state model; and Column 3: the estimated predictive densities of y_1, y_2, and y_3 in a 2-state model. The middle solid curves represent the means, and the upper and lower dashed curves represent the 95%-pointwise confidence intervals.

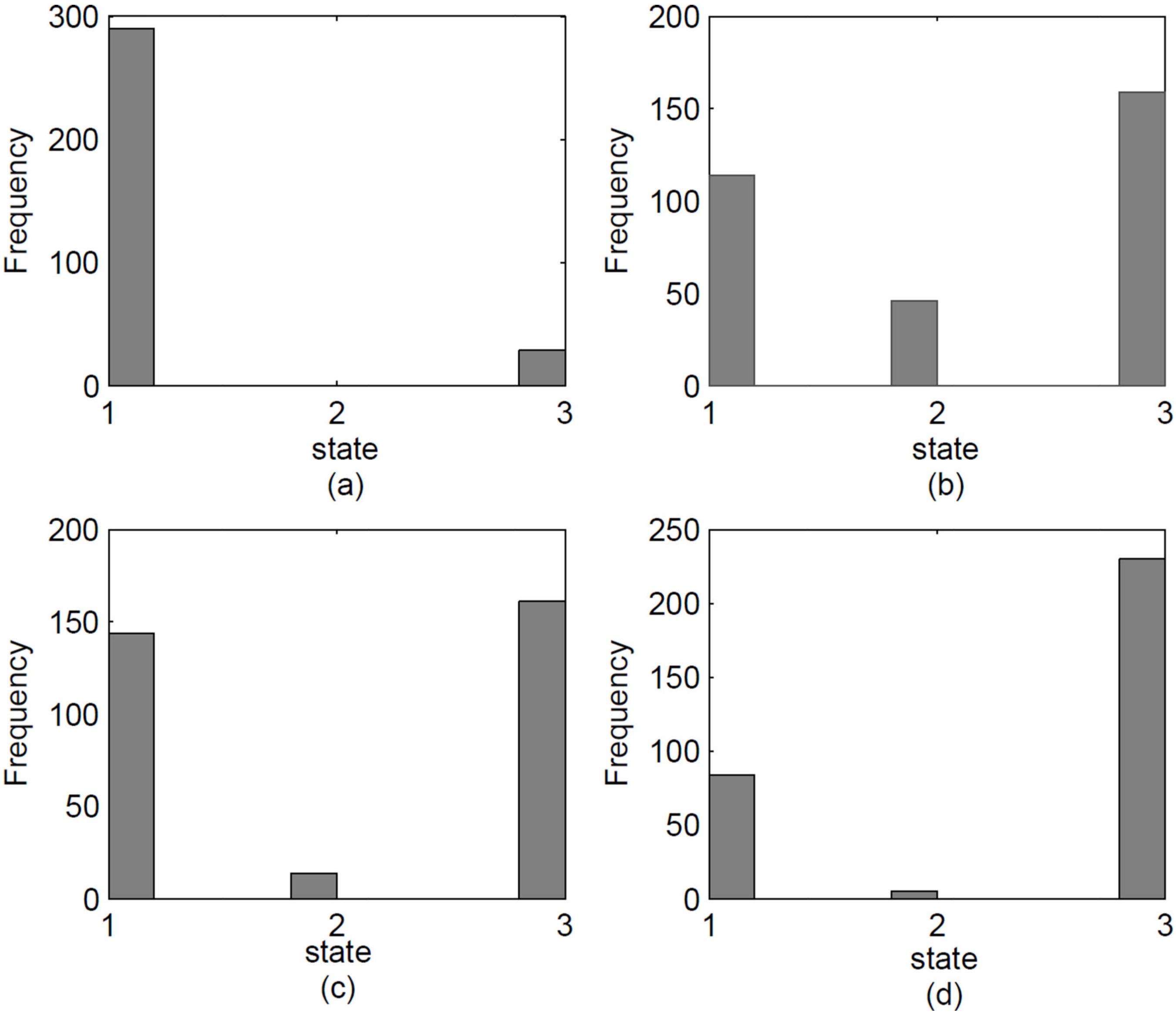

Figure 3.

Estimated frequency of optimal states of patients in cocaine use study: (a) $t = 1$; (b) $t = 2$; (c) $t = 3$; and (d) $t = 4$.

Table 1

Estimates of parameters in Simulation 1 based on 100 replications: $N = 300$, $T = 4$.

	Conditional LVM														
	State 1					State 2					State 3				
Par.	True	BIAS	RMS	SE	CV	True	BIAS	RMS	SE	CV	True	BIAS	RMS	SE	CV
μ_1^s	−0.5	0.000	0.016	0.019	98	0.2	0.006	0.071	0.083	98	0.9	−0.001	0.078	0.092	97
μ_2^s	−0.5	0.003	0.017	0.020	98	0.2	0.005	0.059	0.087	100	0.9	0.005	0.076	0.095	96
μ_3^s	−0.5	−0.001	0.018	0.021	97	0.2	0.013	0.069	0.089	97	0.9	−0.007	0.073	0.096	98
μ_4^s	−0.5	0.000	0.017	0.022	100	0.2	0.002	0.063	0.088	100	0.9	−0.002	0.072	0.094	98
μ_5^s	−0.5	0.003	0.022	0.021	96	0.2	0.000	0.072	0.091	96	0.9	0.008	0.084	0.101	95
μ_6^s	−0.5	0.000	0.015	0.018	96	0.2	0.004	0.075	0.088	96	0.9	−0.007	0.081	0.094	99
μ_7^s	−0.5	0.000	0.018	0.021	95	0.2	0.001	0.060	0.087	96	0.9	−0.001	0.072	0.095	95
μ_8^s	−0.5	0.000	0.017	0.024	99	0.2	0.010	0.065	0.085	100	0.9	0.009	0.075	0.095	100
μ_9^s	−0.5	0.002	0.015	0.021	100	0.2	0.001	0.066	0.089	97	0.9	0.004	0.073	0.095	99
λ_{21}^s	−0.5	0.002	0.031	0.037	94	0.2	−0.002	0.066	0.093	100	0.9	−0.001	0.042	0.058	98
λ_{31}^s	−0.5	0.003	0.034	0.038	98	0.2	−0.012	0.071	0.094	97	0.9	−0.006	0.040	0.055	96
λ_{52}^s	−0.5	0.002	0.018	0.022	97	0.2	0.001	0.072	0.097	98	0.9	−0.005	0.078	0.101	99
λ_{62}^s	−0.5	−0.002	0.017	0.021	99	0.2	0.013	0.076	0.091	96	0.9	0.000	0.080	0.094	99
λ_{83}^s	−0.5	0.002	0.018	0.027	96	0.2	0.007	0.073	0.091	96	0.9	0.006	0.080	0.097	98
λ_{93}^s	−0.5	−0.001	0.016	0.021	98	0.2	−0.014	0.068	0.092	98	0.9	−0.004	0.070	0.092	99

$\psi^s_{\varepsilon 1}$	0.25	0.001	0.013	0.017	96	0.75	−0.001	0.072	0.110	98	1.25	0.017	0.136	0.156	97
$\psi^s_{\varepsilon 2}$	0.25	0.001	0.012	0.015	100	0.75	−0.014	0.087	0.107	97	1.25	0.001	0.096	0.151	99
$\psi^s_{\varepsilon 3}$	0.25	−0.001	0.010	0.013	96	0.75	−0.017	0.082	0.108	98	1.25	0.007	0.116	0.152	98
$\psi^s_{\varepsilon 4}$	0.25	0.004	0.013	0.015	96	0.75	−0.006	0.084	0.110	97	1.25	−0.016	0.124	0.149	97
$\psi^s_{\varepsilon 5}$	0.25	−0.001	0.015	0.018	97	0.75	−0.018	0.082	0.107	99	1.25	−0.027	0.113	0.152	98
$\psi^s_{\varepsilon 6}$	0.25	0.000	0.012	0.015	97	0.75	−0.017	0.089	0.113	95	1.25	−0.012	0.108	0.149	98
$\psi^s_{\varepsilon 7}$	0.25	0.003	0.017	0.012	95	0.75	−0.008	0.074	0.114	95	1.25	0.012	0.122	0.152	93
$\psi^s_{\varepsilon 8}$	0.25	−0.002	0.014	0.014	93	0.75	−0.007	0.078	0.111	94	1.25	−0.009	0.118	0.151	99
$\psi^s_{\varepsilon 9}$	0.25	0.002	0.012	0.015	98	0.75	−0.02	0.078	0.109	97	1.25	−0.003	0.116	0.153	96
Γ^s_1	−0.2	0.000	0.016	0.021	98	0.3	−0.003	0.071	0.090	98	0.8	−0.003	0.094	0.122	95
Γ^s_2	−0.2	−0.001	0.015	0.025	98	0.3	0.001	0.074	0.096	96	0.8	−0.007	0.094	0.120	95
ψ^s_{ζ}	0.25	−0.002	0.012	0.015	98	0.75	−0.021	0.090	0.108	93	1.25	−0.018	0.122	0.150	96
ϕ^s_{11}	1.0	−0.010	0.049	0.058	95	1.0	−0.016	0.109	0.148	92	1.0	0.009	0.102	0.125	99
ϕ^s_{12}	0.0	−0.001	0.034	0.042	97	0.3	−0.008	0.080	0.110	93	0.6	0.010	0.085	0.103	99
ϕ^s_{22}	1.0	−0.008	0.054	0.056	96	1.0	0.001	0.114	0.150	96	1.0	0.003	0.100	0.126	98
Total		- -	0.587	0.701	- -	- -		2.290	3.036	- -	- -		2.720	3.464	- -
Transition model															
Par.		BIAS	RMS	SE	CV	Par		BIAS	RMS	SE	CV				
α_{11}	0.3	0.022	0.086	0.107	100	α_{32}	0.7	0.025	0.313	0.407	99				
α_{12}	0.3	0.041	0.145	0.171	99	β_1	0.7	0.014	0.077	0.092	98				

α_{21}	0.5	0.045	0.245	0.298	97	β_2	0.7	0.014	0.076	0.089	100
α_{22}	0.5	0.067	0.320	0.472	100	σ	1.0	0.038	0.099	0.115	98
α_{31}	0.7	0.016	0.193	0.253	99	Total		- -	1.554	2.004	- -

Table 2

The summary of AIC and BIC in Simulation 2.

		Level of (*N*, *T*)							
		(300 4)		**(1000 4)**		**(300 10)**		**(1000 10)**	
True Model		**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
1-state	1-state	30617	30899	102044	102398	76391	76674	254352	254706
	2-state	31424	32045	110793	111557	76791	77412	256596	257376
		(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	3-state	31462	32412	112067	113261	76829	77779	256599	257792
		(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	4-state	31535	32833	115612	117243	76875	78173	256427	258058
		(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
2-state	1-state	31002	31284	103285	103639	70489	70772	235012	235366
		(97)	(97)	(100)	(100)	(98)	(98)	(100)	(100)
	2-state	26292	26913	87385	88165	63524	64145	211281	211958
	3-state	26882	28180	90721	91914	63567	64518	211557	212337
		(95)	(95)	(100)	(100)	(98)	(98)	(100)	(100)
	4-state	31293	32243	88949	90580	63780	65078	211869	213499
		(98)	(98)	(100)	(100)	(99)	(99)	(100)	(100)
3-state	1-state	32080	32362	106855	107209	76426	76708	255813	256167
		(94)	(94)	(100)	(100)	(100)	(100)	(100)	(100)
	2-state	27960	28562	95758	96538	67182	68131	220938	222131
		(96)	(96)	(100)	(100)	(99)	(99)	(100)	(100)
	3-state	26332	27206	95266	96459	67146	67767	218981	219761
	4-state	25281	26424	95109	96739	68752	69429	222611	224242
		(93)	(93)	(100)	(100)	(98)	(98)	(100)	(100)

Note: The values in parentheses represent correct rates of AIC and BIC in selection of true and posited models.

Table 3

Selection of the transition model in the analysis of cocaine use data.

Transition model		MAR		MNAR	
		AIC	BIC	AIC	BIC
2-state model	without random effect	28251	28514	27290	29797
	with random effect	25579	25847	27826	27558
3-state model	without random effect	19902	20072	20207	20377
	with random effect	17800	17973	18128	18301
4-state model	without random effect	18929	19294	18974	19396
	with random effect	18292	18661	18336	18763

Note: $\psi_{\varepsilon j}^{s}=1.0$.

Table 4

Parameter estimates in cocaine use study.

Para.	Est.	SE	*P*-value	Para.	Est.	SE	*P*-value
μ_1^1	0.213	0.033	<0.001	$\psi_{\varepsilon 4}^1$	0.485	0.011	<0.001
μ_2^1	−0.239	0.028	<0.001	$\psi_{\varepsilon 5}^1$	0.150	0.022	<0.001
μ_3^1	−0.413	0.018	<0.001	$\psi_{\varepsilon 6}^1$	0.410	0.024	<0.001
μ_4^1	0.085	0.015	<0.001	$\psi_{\varepsilon 2}^2$	0.808	0.061	<0.001
μ_5^1	0.140	0.050	0.005	$\psi_{\varepsilon 3}^2$	0.468	0.010	<0.001
μ_6^1	−0.113	0.035	0.001	$\psi_{\varepsilon 4}^2$	0.440	0.046	<0.001
μ_1^2	−0.181	0.012	<0.001	$\psi_{\varepsilon 5}^2$	0.294	0.005	<0.001
μ_2^2	0.570	0.104	<0.001	$\psi_{\varepsilon 6}^2$	0.334	0.008	<0.001
μ_3^2	1.482	0.145	<0.001	$\psi_{\varepsilon 2}^3$	1.022	0.378	0.007
μ_4^2	0.970	0.054	<0.001	$\psi_{\varepsilon 3}^3$	0.795	0.212	<0.001
μ_5^2	1.032	0.076	<0.001	$\psi_{\varepsilon 4}^3$	0.808	0.285	0.005
μ_6^2	0.824	0.041	<0.001	$\psi_{\varepsilon 5}^3$	0.372	0.159	0.019
μ_1^3	−0.110	0.059	0.062	$\psi_{\varepsilon 6}^3$	0.522	0.284	0.066
μ_2^3	0.536	0.122	<0.001	ψ_{ζ}^1	0.933	0.047	<0.001
μ_3^3	0.630	0.165	<0.001	ψ_{ζ}^2	0.885	0.119	<0.001

Para.	Est.	SE	*P*-value	Para.	Est.	SE	*P*-value
μ_4^3	0.433	0.185	0.019	ψ_ζ^3	0.954	0.560	0.088
μ_5^3	0.834	0.394	0.003	ϕ_{11}^1	0.163	0.025	<0.001
μ_6^3	0.571	0.312	0.007	ϕ_{12}^1	−0.023	0.011	0.036
λ_{32}^1	0.432	0.071	<0.001	ϕ_{22}^1	0.278	0.021	0.000
λ_{53}^1	1.839	0.158	<0.001	ϕ_{11}^2	0.388	0.065	<0.001
λ_{63}^1	1.600	0.080	<0.001	ϕ_{12}^2	−0.067	0.126	<0.001
λ_{32}^2	0.320	0.117	0.006	ϕ_{22}^2	0.178	0.075	0.017
λ_{53}^2	1.715	0.212	<0.001	ϕ_{11}^3	0.206	0.096	0.032
λ_{63}^2	1.614	0.184	<0.001	ϕ_{12}^3	0.138	0.026	<0.001
λ_{32}^3	0.079	0.037	0.033	ϕ_{22}^3	0.382	0.131	0.004
λ_{53}^3	1.031	0.112	<0.001	π_1	0.854	0.256	0.001
λ_{63}^3	0.816	0.110	<0.001	π_2	0.145	0.057	0.011
Γ_1^1	−0.782	0.249	0.002	γ_{11}	6.708	0.568	<0.001
Γ_2^1	0.132	0.014	<0.001	γ_{12}	6.133	0.264	<0.001
Γ_1^2	−0.217	0.115	0.059	γ_{21}	4.169	0.350	<0.001
Γ_2^2	0.343	0.073	<0.001	γ_{22}	3.587	0.154	<0.001
Γ_1^3	−0.164	0.078	0.036	γ_{31}	3.111	0.699	<0.001

Para.	Est.	SE	*P*-value	Para.	Est.	SE	*P*-value
Γ_2^3	0.573	0.167	0.001	γ_{32}	3.584	0.615	<0.001
$\psi_{\varepsilon 2}^1$	0.524	0.006	<0.001	β_1	0.374	0.185	0.043
$\psi_{\varepsilon 3}^1$	0.308	0.003	<0.001	σ	0.891	0.128	<0.001

Note: λ_{32}^s, λ_{53}^s, and λ_{63}^s are state-specific factor loadings for the 2nd indicator of treatment, and the 2nd and 3rd indicators of psychological problem; Γ_1^s and Γ_2^2 are state-specific effects of treatment and psychological problem on cocaine use; αs are intercepts from the continuation ratio model; β is the effect of patients' current employment on the transition probability; and σ is the variance of the random effect in the continuation ratio model.