

Published in final edited form as:

Stat Med. 2008 July 20; 27(16): 3017–3041. doi:10.1002/sim.3266.

A two-level structural equation model approach for analyzing multivariate longitudinal responses

Xin-Yuan Song^{1,*,†,‡}, Sik-Yum Lee¹, and Yih-Ing Hser²

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

²Integrated Substance Abuse Programs, University of California, Los Angeles, CA, U.S.A.

SUMMARY

The analysis of longitudinal data to study changes in variables measured repeatedly over time has received considerable attention in many fields. This paper proposes a two-level structural equation model for analyzing multivariate longitudinal responses that are mixed continuous and ordered categorical variables. The first-level model is defined for measures taken at each time point nested within individuals for investigating their characteristics that are changed with time. The second level is defined for individuals to assess their characteristics that are invariant with time. The proposed model accommodates fixed covariates, nonlinear terms of the latent variables, and missing data. A maximum likelihood (ML) approach is developed for the estimation of parameters and model comparison. Results of a simulation study indicate that the performance of the ML estimation is satisfactory. The proposed methodology is applied to a longitudinal study concerning cocaine use.

Keywords

latent variables; longitudinal study on cocaine use; maximum likelihood; MCEM algorithm; model comparison; ordered categorical variables

1. INTRODUCTION

In public health and biomedical sciences, it is common to encounter latent variables or constructs that cannot be directly measured by a single observed variable, but instead assessed through a number of observed variables. Structural equation models (SEMs) [1] are a flexible class of models for complex modeling of multivariate correlated data to explain the interrelationships among observed and latent variables. In general, SEMs combine the ideas of factor analysis and regression. They are formulated by two major components. The first component is the measurement equation, which is basically a confirmatory factor analysis model for grouping correlated observed variables to 'measure' their corresponding latent variables (factors) and for taking the measurement errors into account. The second component is a regression-type structural equation for assessing the effects of independent latent variables on dependent latent variables of interest. As the number of latent variables is much less than the number of original observed variables, the use of a structural equation with latent variables has its advantages over the application of the ordinary regression model with originally observed variables. Through the use of user-friendly software [2,3], SEMs have been

Copyright © 2008 John Wiley & Sons, Ltd.

*Correspondence to: Xin-Yuan Song, Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong..

†xysong@sta.cuhk.edu.hk

‡Assistant Professor.

extensively applied to behavioral, educational, and social–psychological sciences in the past years and to biological and medical sciences in recent years. The basic theory and statistical properties of SEMs are mainly developed in the field of psychometrics for assessing latent constructs. Recently, advanced SEM methodologies have also appeared in mainstream statistical journals; see, for example, [4–10]. Moreover, Sanchez *et al.* [11] presented a review of the basic theory, including its connection to some latent variables models.

In this paper, we propose a novel, dynamic two-level SEM to analyze multivariate response variables measured at multiple time points. More specifically, let \mathbf{u}_{gt} be a random vector for the g th ($g=1, \dots, G$) individual measured at time point t ($t=1, \dots, T$). A two-level model for \mathbf{u}_{gt} will be defined by $\mathbf{u}_{gt} = \mathbf{y}_g + \mathbf{v}_{gt}$, where \mathbf{y}_g is the second-level random vector that is independent of t , and \mathbf{v}_{gt} is the first-level random vector for accounting characteristics that change dynamically with t . A confirmatory factor analysis model with fixed covariates will be defined for \mathbf{y}_g for accounting characteristics that are invariant with t . A nonlinear SEM with fixed covariates will be defined for \mathbf{v}_{gt} , in which the latent variables at every time point are divided into independent and dependent latent variables, and their dynamic relationships over time are assessed by a flexible autoregressive nonlinear structural equation (see equation (5)). Here, latent variables are used to represent latent traits that are related to several observed variables (indicators). To cope with the complex nature of real-world data in substantive research, the proposed model also accommodates mixed continuous and ordered categorical variables, and missing data that are missing at random. A maximum likelihood (ML) approach will be developed for analyzing the proposed model. A Monte Carlo Expectation–Maximization (MCEM) algorithm [12] will be developed for obtaining the ML estimates, and the Louis formula [13] will be used to obtain the standard error estimates. The Bayesian information criterion (BIC) will be used for model comparison. Owing to the complexity of the model, the observed-data likelihood involved in the BIC includes intractable multiple integrals. A procedure based on path sampling [14] will be developed to evaluate the observed-data likelihood.

We illustrate our methodology through a longitudinal data set about cocaine use and related phenomena. Cocaine use is a major social and health problem in the United States. According to the 2002 National Survey on Drug Use and Health [15], an estimated 34 million Americans aged 12 years or older (14.4 per cent) reported cocaine use at least once in their lifetime; among them, an estimated 2 million persons reported cocaine use in the past month. Cocaine is consistently detected in the urine specimens from approximately one-third of arrestees tested across the nation each year [16]. Furthermore, cocaine is the illegal drug mentioned most often in emergency department records [17]. There are various latent traits that influence cocaine use. Specifically, many studies have shown substantial impact of psychiatric problems on cocaine-dependent patients [18–20] and that social support is another key latent trait [21–23].

Although treatment for cocaine use has received considerable attention in recent years [24, 25], the longitudinal phenomena of cocaine use in relation to psychiatric problems and social support have yet to be adequately explicated. To illustrate the application of the proposed methodologies, we analyze the UCLA longitudinal data set collected from cocaine-dependent patients at intake, 1 year, 2 years, and 12 years after treatment [26]. As this data set involves ordered categorical outcomes with missing entries, the proposed two-level SEM is required to assess the longitudinal effects of the latent traits (psychiatric problems and social support) to cocaine use.

Although some specific SEMs or latent variable models have been developed for analyzing longitudinal data, their objectives and/or formulations are quite different from the proposed dynamic two-level nonlinear SEMs. For instance, a standard growth curve model [27,28] can be viewed as the following confirmatory factor analysis model: $\mathbf{x}_i = \mathbf{A}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$, where \mathbf{x}_i is a

vector of repeated measures of a univariate response variable, Λ is the factor loading matrix of sequential known values of the growth curve records, and ω_i is the latent growth factor containing the 'initial status (intercept)' and 'rate of change (slope)', and ϵ_i is the vector of residuals. This is a single-level linear SEM, in which the latent variables in ω_i are used to assess the characteristic of change over time rather than latent traits such as 'psychiatric problem' and 'social support' in our example. Hence, the formulation of a growth curve model is less general, and the interpretation of the latent variables is very different. Dunson [29] recently analyzed multidimensional longitudinal data by developing dynamic latent variable models, which allow mixtures of count, and categorical and continuous response variables. In this model, to assess the changes between the latent variables at different time points, the latent variable ω_{it} at time t for individual i was regressed on the linear terms of the past latent vectors $\omega_{i1}, \dots, \omega_{i,t-1}$ through a regression equation with fixed covariates. Note that Dunson's [29] model is single-level and its latent variables are not divided into independent and dependent latent variables. Compared with Dunson's single-level model, our two-level model includes a second-level model for assessing the individuals' characteristics that are invariant with time. Moreover, latent variables at the first level of our model are divided into independent and dependent latent variables, and the dynamic nonlinear effects of the independent latent variables on the dependent latent variables are assessed through a rather general autoregressive nonlinear structural equation (see equation (5)). Similarly, the multilevel model [30] for the longitudinal profiling of health-care units does not involve an autoregressive nonlinear structural equation among covariates and latent variables. Other longitudinal models in the statistics literature, such as the linear mixed model [31] and the generalized linear mixed model [32], do not involve the latent traits and their associated structural equation. Clearly, the objectives and formulations of these models and our proposed model are very different. Finally, refer to the second remark given in Section 2.3 for the differences of the proposed model and the existing multilevel SEMs.

The contributions of this paper include (i) a novel dynamic two-level nonlinear SEM with fixed covariates for analyzing longitudinal data from mixed continuous and ordered categorical measures and missing data, (ii) a novel utilization of the computational tools, such as the MCEM algorithm [12] and path sampling [14] with the derivation of some new conditional distributions and derivatives for estimation and model comparison of the proposed model, and (iii) a novel application of the developed methodologies to the longitudinal study of cocaine use.

Section 2 presents the two-level SEM for analyzing multivariate random responses with mixed continuous and ordered categorical variables measured at multiple time points. An ML approach for estimation and model comparison in the proposed model framework is discussed in Section 3. A simulation study and a real example using the longitudinal data of cocaine use are presented in Sections 4 and 5, respectively. In Section 6, we discuss some limitations and outline several extensions for further research.

2. THE SEM FOR MULTIVARIATE LONGITUDINAL DATA

Consider a set of observations of a $p \times 1$ random vector \mathbf{u}_{gt} for individual $g = 1, \dots, G$, which was measured at multiple time points $t = 1, \dots, T$. The collection $\mathbf{U} = \{\mathbf{u}_{gt} : g = 1, \dots, G, t = 1, \dots, T\}$ can be regarded as an observed sample of hierarchical observations that were measured at different time points (first level) and were nested in the individuals (second level). Hence, the following two-level SEM is proposed to model \mathbf{u}_{gt} :

$$\mathbf{u}_{gt} = \mathbf{y}_g + \mathbf{v}_{gt} \quad (1)$$

where \mathbf{y}_g and \mathbf{v}_{gt} are the random vectors that, respectively, model the second-level effect of the individual g and the first-level effect with respect to the individual g at time t .

2.1. The second-level model

The second-level model for assessing characteristics of individuals that are invariant over time is defined by

$$\mathbf{y}_g = \mathbf{A}_0 \mathbf{c}_{g0} + \Lambda_0 \omega_{g0} + \varepsilon_{g0} \tag{2}$$

where \mathbf{c}_{g0} is a vector of fixed covariates, ω_{g0} is a vector of latent variables, ε_{g0} is a vector of residual errors, and \mathbf{A}_0 and Λ_0 are matrices of unknown coefficients. We assume that ω_{g0} is identically and independently distributed (i.i.d.) as $N[\mathbf{0}, \Phi_0]$ and that ε_{g0} is independent of ω_{g0} , and i.i.d. as $N[\mathbf{0}, \Psi_0]$, where Ψ_0 is a diagonal matrix. This is a factor analysis model with covariates [33], which is defined for studying the relationships between the observed variables in \mathbf{u}_{gt} and the latent variables in ω_{g0} with respect to different individuals but common to all time points.

2.2. The first-level model

We define the following measurement model for the more important first-level random vector \mathbf{v}_{gt} :

$$\mathbf{v}_{gt} = \mathbf{A}_t \mathbf{c}_{gt} + \Lambda_t \omega_{gt} + \varepsilon_{gt}, \quad g=1, \dots, G, \quad t=1, \dots, T \tag{3}$$

in which the definitions of \mathbf{A}_t , \mathbf{c}_{gt} , Λ_t , ω_{gt} , and ε_{gt} are similar to those given in equation (2); except here they are defined at time point t nested within the individual g , and the distribution of ε_{gt} is $N[\mathbf{0}, \Psi_t]$, where Ψ_t is assumed to be a diagonal matrix for brevity. Hence, we assume that the random vector \mathbf{u}_{gt} conditional on \mathbf{y}_g has the following structure:

$$\mathbf{u}_{gt} = \mathbf{y}_g + \mathbf{A}_t \mathbf{c}_{gt} + \Lambda_t \omega_{gt} + \varepsilon_{gt}, \quad g=1, \dots, G, \quad t=1, \dots, T \tag{4}$$

Conditional on \mathbf{y}_g , equation (4) accounts for dependency among the observed variables measured for the individual g at a given time point t through the fixed covariates \mathbf{c}_{gt} and the shared latent variables ω_{gt} . The relationships between the observed variables and the corresponding fixed covariates and latent variables that change dynamically over time can be assessed by estimating \mathbf{A}_t and Λ_t .

According to the spirit of SEM, we consider a partition of ω_{gt} into $(\eta'_{gt}, \xi'_{gt})'$, where η_{gt} and ξ_{gt} are q_1 -dimensional-dependent and q_2 -dimensional-independent random latent vectors, respectively, $q=q_1+q_2$. Effects of the independent latent variables ξ_{gt} on the dependent latent variables η_{gt} are studied through the following autoregressive nonlinear structural equation: For $g=1, \dots, G, t=1, \dots, T$,

$$\eta_{gt} = \mathbf{B}_0 \mathbf{d}_{g0} + \mathbf{B}_t \mathbf{d}_{gt} + \Gamma_t^* \mathbf{F}_t^* (\omega_{g1}, \dots, \omega_{g,t-1}, \xi_{gt}) + \delta_{gt} \tag{5}$$

where \mathbf{B}_0 , \mathbf{B}_t , and Γ_t^* are matrices of unknown coefficients, \mathbf{d}_{g0} is a vector of fixed covariates, which is independent of t , \mathbf{d}_{gt} is another vector of fixed covariates, δ_{gt} is a vector of residual errors, and $\mathbf{F}_t^* = (f_{t1}, \dots, f_{tr})$ is a vector-valued function in which f_{tj} is a differentiable function of the independent latent vector ξ_{gt} at the current time point and the latent vectors $\omega_{g1}, \dots, \omega_{g,t-1}$ at previous time points. The residual vector δ_{gt} is independent of ε_{gt} , $\omega_{g1}, \dots, \omega_{g,t-1}$, and ξ_{gt} . Moreover, these residual errors are i.i.d. $N[\mathbf{0}, \Psi_{\delta t}]$, where $\Psi_{\delta t}$ is a diagonal matrix. Let

$\xi_g = (\xi'_{g1}, \dots, \xi'_{gt})'$; the distribution of ξ_g is assumed to be $N[\mathbf{0}, \Phi]$, where Φ contains the variance-covariance matrix Φ_{tt} of ξ_{gt} , and the covariance matrix Φ_{tjt_i} of ξ_{gt_i} and ξ_{gt_j} at different time points t_i and t_j .

2.3. Remarks

1. The first and second terms of equation (5) allow for direct effects of observed predictors that are invariant over time (\mathbf{d}_{g0}) and variant over time (\mathbf{d}_{gt}) on the dependent latent variables. The term $\Gamma_t^* \mathbf{F}_t^*(\cdot)$ is very important for assessing the effects of independent latent variables on dependent latent variables to change dynamically over time. It allows various flexible autoregressive structures for recovering the dependency of the dependent latent variable η_{gt} at the current time point and across both dependent latent variables at previous time points and independent latent variables at previous and current time points. For example, typical special cases are

$$\begin{aligned} \text{(a) } \eta_{gt} &= \beta_{t-1} \eta_{g,t-1} \\ &+ \Gamma_1 \xi_{g1} + \dots + \Gamma_{t-1} \xi_{g,t-1} \\ &+ \Gamma_t \mathbf{F}_t(\xi_{gt}) + \delta_{gt} \end{aligned}$$

$$\begin{aligned} \text{(b) } \eta_{gt} &= \beta_1 \eta_{g1} + \dots + \beta_{t-1} \eta_{g,t-1} \\ &+ \Gamma_t \mathbf{F}_t(\xi_{gt}) + \delta_{gt} \end{aligned}$$

where \mathbf{F}_t is a vector of differentiable functions of ξ_{gt} . In the special case (a), η_{gt} is assessed by the effects of linear-independent latent vectors at all previous time points, the dependent-latent vector at $t-1$, and finally nonlinear terms of independent latent variables at the current time point. The autoregressive structure (b) allows the dependency of η_{gt} on the previous endogenous latent vector $\eta_{g1}, \dots, \eta_{g,t-1}$ and a general function $\mathbf{F}_t(\xi_{gt})$ of the current exogenous latent vector ξ_{gt} . Note that the longitudinal models developed in Dunson [29] and Daniels and Normand [30] do not involve a structural equation in relation to covariates and latent variables.

2. Other multilevel SEMs have been developed to analyze hierarchical data collected from units that are nested within clusters [8,34]. Owing to different objectives, the formulations of the first-level model for \mathbf{v}_{gt} in those multilevel SEMs are quite different and do not accommodate certain important features of our dynamic model. For example, their matrices of coefficient parameters and covariance matrices of errors are variant with cluster g rather than t . Moreover, for $t \neq r$, their \mathbf{v}_{gt} and \mathbf{v}_{gr} are assumed to be independent; however, \mathbf{v}_{gt} and \mathbf{v}_{gr} could be correlated in our model.

2.4. Ordered categorical data and identification of the model

From its definition, \mathbf{u}_{gt} is a vector of continuous observations. However, in most substantive research, the continuous measurements of some components of \mathbf{u}_{gt} may not be available. Suppose that \mathbf{u}_{gt} is composed of a subvector \mathbf{x}_{gt} of observed measurements and a subvector \mathbf{w}_{gt} of unobserved measurements whose information is given by ordered categorical observations. In the generic sense, the relationship between an ordered categorical variable z and its underlying continuous variable w is defined by $z=k+1$ if $\alpha_k \leq w < \alpha_{k+1}$, for $k=0, \dots, m-1$, where $\{-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_m = \infty\}$ is the set of threshold parameters that define the m categories. A special case with $m=2$ is the dichotomous or ordered binary variable. Missing data that are missing at random (MAR, [35]) are handled using the procedure given in [36]. Details are not included here for brevity.

The proposed model is not identified without imposing identification conditions. For example, the variance and thresholds associated with each ordered categorical variable are not identifiable; the first- and second-level covariance structures are not identified, and we cannot allow intercepts to simultaneously exist in equations (3) and (5). Existing methods suggested in the SEM literature can be adopted for identifying various components of the current model. For example, the identification problem associated with an ordered categorical variable can be solved by fixing the thresholds α_1 and/or α_{m-1} at some appropriate preassigned values [6, 36], and the covariance structures in the first- and second-level models can be identified by the common practice in structural equation modeling by fixing appropriate elements in Λ_0 and Λ_t at preassigned values [8].

3. AN ML APPROACH

3.1. Estimation

Let α be the parameter vector of the unknown thresholds, and θ be the parameter vector that contains all the unknown distinct structural parameters that are involved in equations (2), (3), and (5). The likelihood function of the observed data involves high-dimensional intractable integrals that are induced by the ordered categorical variables and latent variables. One approach to handle this integral is using adaptive quadrature [34]. In this paper, we use the idea of data augmentation and the MCEM algorithm [12] in the ML estimation.

Let $\mathbf{z}_{gt} = \{z_{gtj}; j = 1, \dots, s\}$ be a vector of ordered categorical observations, and let $\mathbf{u}_{gt} = \{\mathbf{w}_{gt}, \mathbf{x}_{gt}\}$, where \mathbf{x}_{gt} is a vector of continuous measurements and \mathbf{w}_{gt} is the latent continuous measurement corresponding to \mathbf{z}_{gt} , $\mathbf{u}_g = \{\mathbf{u}_{gt}; t = 1, \dots, T\}$, $\mathbf{Z} = \{\mathbf{z}_{gt}; g = 1, \dots, G, t = 1, \dots, T\}$, $\mathbf{W} = \{\mathbf{w}_{gt}; g = 1, \dots, G, t = 1, \dots, T\}$, and $\mathbf{X} = \{\mathbf{x}_{gt}; g = 1, \dots, G, t = 1, \dots, T\}$. Moreover, let $\omega_g = (\eta'_{g1}, \dots, \eta'_{gT}, \xi'_{g1}, \dots, \xi'_{gT})'$, $\Omega_1 = \{\omega_g; g = 1, \dots, G\}$, $\Omega_0 = \{\omega_{g0}; g = 1, \dots, G\}$, and $\mathbf{Y} = \{\mathbf{y}_g; g = 1, \dots, G\}$ be matrices of latent vector at the first and second levels. Further, let $\Pi_t = (\mathbf{B}'_t \Gamma_t^*)'$, and $\mathbf{h}_{gt} = (\mathbf{d}'_{gt}, \mathbf{F}_t^* (\omega_{g1}, \dots, \omega_{g,t-1}, \xi_{gt})')$, then equation (5) can be rewritten as

$$\eta_{gt} = \mathbf{B}_0 \mathbf{d}_{g0} + \Pi_t \mathbf{h}_{gt} + \delta_{gt} \tag{6}$$

The observed data set is (\mathbf{X}, \mathbf{Z}) . The observed-data likelihood is very complicated due to the existence of latent quantities $(\mathbf{Y}, \Omega_1, \Omega_2, \mathbf{W})$. Utilizing the idea of data augmentation, we consider the complete data set, which is equal to $\{\mathbf{Y}, \Omega_0, \Omega_1, \mathbf{W}, \mathbf{X}, \mathbf{Z}\} = \{\mathbf{Y}, \Omega_0, \Omega_1, \mathbf{U}, \mathbf{Z}\}$. The complete data likelihood function is equal to

$$\begin{aligned} p & (\mathbf{Y}, \Omega_0, \Omega_1, \mathbf{U}, \mathbf{Z} | \alpha, \theta) \\ & = p(\mathbf{Z} | \mathbf{Y}, \Omega_0, \Omega_1, \mathbf{U}, \alpha, \theta) p(\mathbf{U} | \mathbf{Y}, \Omega_1, \theta) p(\Omega_1 | \theta) p(\mathbf{Y} | \Omega_0, \theta) p(\Omega_0 | \theta) \\ & = (2\pi)^{-GT_p/2} |\Psi_t|^{-GT/2} \prod_{g=1}^G \prod_{t=1}^T \exp \left\{ -\frac{1}{2} (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt})' \Psi_t^{-1} \right. \\ & \quad \times \left. (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt}) \right\} I_{\mathbf{R}_{gt}}(\mathbf{w}_{gt}) \\ & \quad \times (2\pi)^{-GT_{q1}/2} |\Psi_{\delta t}|^{-GT/2} \prod_{g=1}^G \prod_{t=1}^T \exp \left\{ -\frac{1}{2} (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt})' \Psi_{\delta t}^{-1} (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt}) \right\} \\ & \quad \times (2\pi)^{-G_{\eta 2}/2} |\Phi|^{-G/2} \prod_{g=1}^G \exp \left\{ -\frac{1}{2} \xi'_g \Phi^{-1} \xi_g \right\} \\ & \quad \times (2\pi)^{-G_{\omega 0}/2} |\Phi_0|^{-G/2} \prod_{g=1}^G \exp \left\{ -\frac{1}{2} \omega'_{g0} \Phi_0^{-1} \omega_{g0} \right\} \\ & \quad \times (2\pi)^{-G_{\psi}/2} |\Psi_0|^{-G/2} \prod_{g=1}^G \exp \left\{ -\frac{1}{2} (\mathbf{y}_g - \mathbf{A}_0 \mathbf{c}_{g0} - \Lambda_0 \omega_{g0})' \Psi_0^{-1} (\mathbf{y}_g - \mathbf{A}_0 \mathbf{c}_{g0} - \Lambda_0 \omega_{g0}) \right\} \end{aligned} \tag{7}$$

where $I_{\mathbf{R}_{gt}}(\mathbf{w}_{gt})$ is an indicator function that takes the value 1 if $\mathbf{w}_{gt} \in \mathbf{R}_{gt}$ and zero otherwise, and $\mathbf{R}_{gt} = [\alpha_{1,z_{gt1}}, \alpha_{1,z_{gt1}+1}] \times \dots \times [\alpha_{s,z_{gts}}, \alpha_{s,z_{gts}+1}]$. Note that for every component w_{gtj} in \mathbf{w}_{gt} , there exists one and only one $[\alpha_{j,z_{gtj}}, \alpha_{j,z_{gtj}+1}]$ such that w_{gtj} is in $[\alpha_{j,z_{gtj}}, \alpha_{j,z_{gtj}+1}]$. Hence, the indicator function and the corresponding value of the density function are nonzero. It can be shown from (7) that the complete-data log-likelihood function can be expressed as $L_c(\mathbf{Y}, \mathbf{\Omega}_0, \mathbf{\Omega}_1, \mathbf{U}, \mathbf{Z}; \mathbf{a}, \boldsymbol{\theta}) = L_1 + L_2$, where

$$L_1 = -\frac{1}{2} \sum_{g=1}^G \left[q_2 \log(2\pi) + \log|\Phi| + \xi'_g \Phi^{-1} \xi_g + \sum_{t=1}^T \{(p+q_1) \log(2\pi) + \log|\Psi_t| + \log|\Psi_{\delta t}| + (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt})' \Psi_t^{-1} (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt}) + (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt})' \Psi_{\delta t}^{-1} (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt}) \} \right] \text{ for } \mathbf{w}_{gt} \in \mathbf{R}_{gt} \tag{8}$$

$$L_2 = -\frac{1}{2} \sum_{g=1}^G \left\{ (p+q) \log(2\pi) + \log|\Psi_0| + \log|\Phi_0| + \omega'_{g0} \Phi_0^{-1} \omega_{g0} + (\mathbf{y}_g - \mathbf{A}_0 \mathbf{c}_{g0} - \Lambda_0 \omega_{g0})' \Psi_0^{-1} (\mathbf{y}_g - \mathbf{A}_0 \mathbf{c}_{g0} - \Lambda_0 \omega_{g0}) \right\} \tag{9}$$

Note that L_c decomposes into two separable functions, each with distinct sets of parameters. This property simplifies the estimation significantly, although due to the complex function \mathbf{h}_{gt} , L_1 is still rather complicated. ML estimates of \mathbf{a} and $\boldsymbol{\theta}$ are obtained using the EM algorithm that consists of the following E-step and M-step at the l th iteration: E-step: Evaluate $Q(\mathbf{a}, \boldsymbol{\theta} | \boldsymbol{\alpha}^{(l)}, \boldsymbol{\theta}^{(l)}) = E\{L_c(\mathbf{Y}, \mathbf{\Omega}_0, \mathbf{\Omega}_1, \mathbf{U}, \mathbf{Z}; \mathbf{a}, \boldsymbol{\theta}) | \mathbf{X}, \mathbf{Z}, \boldsymbol{\alpha}^{(l)}, \boldsymbol{\theta}^{(l)}\}$, in which the expectation is taken with respect to the conditional distribution of $(\mathbf{Y}, \mathbf{\Omega}_0, \mathbf{\Omega}_1, \mathbf{W})$ given (\mathbf{X}, \mathbf{Z}) at $(\boldsymbol{\alpha}^{(l)}, \boldsymbol{\theta}^{(l)})$. M-step: Determine $(\boldsymbol{\alpha}^{(l+1)}, \boldsymbol{\theta}^{(l+1)})$ by maximizing $Q(\mathbf{a}, \boldsymbol{\theta} | \boldsymbol{\alpha}^{(l)}, \boldsymbol{\theta}^{(l)})$.

Evaluation of the E-step is rather complicated because the conditional expectations involve intractable multiple integrals. Inspired by the MCEM algorithm [12], we solve this problem by simulating a sufficiently large number of observations from the corresponding conditional distributions using a hybrid algorithm that combines the Gibbs sampler [37] and the Metropolis–Hastings (MH) algorithm [38,39]. In this algorithm, observations are sampled interactively from the following conditional distributions: $p(\mathbf{Y} | \mathbf{\Omega}_0, \mathbf{\Omega}_1, \mathbf{U}, \mathbf{Z}, \mathbf{a}, \boldsymbol{\theta})$, $p(\mathbf{\Omega}_0 | \mathbf{Y}, \mathbf{\Omega}_1, \mathbf{U}, \mathbf{Z}, \mathbf{a}, \boldsymbol{\theta})$, $p(\mathbf{\Omega}_1 | \mathbf{Y}, \mathbf{\Omega}_0, \mathbf{U}, \mathbf{Z}, \mathbf{a}, \boldsymbol{\theta})$, and $p(\mathbf{W} | \mathbf{Y}, \mathbf{\Omega}_0, \mathbf{\Omega}_1, \mathbf{X}, \mathbf{Z}, \mathbf{a}, \boldsymbol{\theta})$. Expressions for these conditional distributions are provided in Appendix A.

The M-step updates unknown parameters by maximizing the conditional expectations obtained in the E-step. Structural parameter $\boldsymbol{\theta}$ can be updated by maximizing the conditional expectation of L_1 and L_2 with respect to $\boldsymbol{\theta}$ by solving the following system of equations:

$$\frac{\partial Q(\mathbf{a}, \boldsymbol{\theta} | \boldsymbol{\alpha}^{(l)}, \boldsymbol{\theta}^{(l)})}{\partial \boldsymbol{\theta}} = E \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} (L_1 + L_2) | \mathbf{X}, \mathbf{Z}, \boldsymbol{\alpha}^{(l)}, \boldsymbol{\theta}^{(l)} \right\} = 0 \tag{10}$$

Threshold parameters \mathbf{a} can be updated according to the method described in Shi and Lee [6], and Lee and Song [8]. The M-step is completed by conditional maximization. Technical details on the derivatives of $Q(\mathbf{a}, \boldsymbol{\theta} | \boldsymbol{\alpha}^{(l)}, \boldsymbol{\theta}^{(l)})$ with respect to components of $\boldsymbol{\theta}$ and the corresponding solutions are given in Appendix B. The convergence of the MCEM algorithm is monitored by the following method proposed in Shi and Copas [40]. After the m_0 th MCEM iteration,

$\theta^{-(l)} = (\theta^{(l-m_0+1)} + \dots + \theta^{(l)}) / m_0$ is computed, and the convergence is monitored using the following stopping rule: For given small values δ_1^* and δ_2^* (e.g. 0.001), the procedure is stopped if $\|\theta^{-(l)} - \theta^{-(l-\gamma_0^*)}\| / (\|\theta^{-(l-\gamma_0^*)}\| + \delta_1^*)$ is smaller than some predetermined small value δ_2^* . To avoid the danger of stopping early, a value of γ_0^* larger than 2 is suggested. Convergence is claimed after the stopping rule is satisfied for several consecutive iterations, and $\theta^{-(l)}$ is then taken to be the ML estimate. The sample observations of Ω_0 and Ω_1 simulated at the last iteration of the MCEM algorithm provide estimates of the latent variables. Shi and Copas [40] argued that for a sufficiently large m_0 , the average of the associative Monte Carlo errors is negligible. To reduce the bias, an appropriate m_0 that can control the Monte Carlo errors within a bearable limit is taken. In the simulation study and the example given in Sections 4 and 5, we take $m_0 = 50$. Finally, good starting values of some parameters (such as Λ_t and Π_t) could be obtained from the estimates achieved through separate analysis of the individual model at each time point. As each individual model only involves a comparatively small number of parameters, little computing effort is required.

Standard error estimates of the structural parameters in θ are obtained by the following Louis formula [24] through the simulated observations and the ML estimates:

$$\begin{aligned} & - \frac{\partial^2 L(\mathbf{X}, \mathbf{Z} | \alpha, \theta)}{\partial \theta \partial \theta^T} \\ & = E \left\{ - \frac{\partial^2 L(\mathbf{Y}, \Omega_0, \Omega_1, \mathbf{U}, \mathbf{Z} | \alpha, \theta)}{\partial \theta \partial \theta^T} \right\} \\ & \quad - \text{Var} \left\{ \frac{\partial L(\mathbf{Y}, \Omega_0, \Omega_1, \mathbf{U}, \mathbf{Z} | \alpha, \theta)}{\partial \theta} \right\} \end{aligned}$$

where expectations are taken with respect to the conditional distribution of $(\mathbf{Y}, \Omega_0, \Omega_1, \mathbf{W})$ given (\mathbf{X}, \mathbf{Z}) and (α, θ) , and the whole expression is evaluated at $(\widehat{\alpha}, \widehat{\theta})$. These expectations are difficult to evaluate analytically due to existence of the latent quantities, but they can be approximated, respectively, by the sample mean and the sample covariance matrix of the random sample $\{(Y^{(j)}, \Omega_0^{(j)}, \Omega_1^{(j)}, \mathbf{W}^{(j)}); j=1, \dots, J\}$, generated from $p(\mathbf{Y}, \Omega_0, \Omega_1, \mathbf{W} | \mathbf{X}, \mathbf{Z}, \widehat{\alpha}, \widehat{\theta})$ using the proposed hybrid algorithm. Details are not given to save space.

3.2. Model comparison

Model comparison is an important issue in SEM analysis. For the current two-level SEM, it is interesting to evaluate competing models that correspond to various hypotheses about the characteristics of the parameters with respect to time changes. In this article, competing models M_1 and M_2 are compared with the following BIC:

$$\begin{aligned} \text{BIC}_{12} = & -2 \left\{ \log p(\mathbf{X}, \mathbf{Z} | \widehat{\alpha}_1, \widehat{\theta}_1, M_1) - \log p(\mathbf{X}, \mathbf{Z} | \widehat{\alpha}_2, \widehat{\theta}_2, M_2) \right\} \\ & + (d_1 - d_2) \log(GT) \end{aligned} \tag{11}$$

where $p(\mathbf{X}, \mathbf{Z} | \widehat{\alpha}_k, \widehat{\theta}_k, M_k)$ is the observed-data likelihood evaluated at the ML estimates of α_k and θ_k under M_k , and d_k is the dimension of θ_k . The criterion given in Kass and Raftery [41] can be used for interpreting BIC_{12} .

In obtaining BIC_{12} , we must compute the observed-data log-likelihood $\log p(\mathbf{X}, \mathbf{Z} | \widehat{\alpha}_h, \widehat{\theta}_h, M_h)$ which involves complicated multiple integrals. Inspired by its successful application in computing the observed-data likelihood for complex latent variable models [8,42], we apply path sampling [14] to compute BIC_{12} . Path sampling is a generalization of bridge sampling [43] and gives more accurate results in computing complicated multiple integrals. It is conceptually simple, easy to implement, and applicable to a wide range of models [14].

4. A SIMULATION STUDY

We have conducted an extensive simulation study in different settings to empirically evaluate the performance of the MCEM algorithm. To save space, only the results obtained under the following settings are reported here. A special case of the two-level SEM defined in (2), (4), and (5) with $p=9$ and $G=500$ at three time points is considered. For the second-level model associated with individuals, \mathbf{c}_{g0} is a 2×1 vector of covariates, in which the first entry is taken to be 1 to represent the intercept, and the second entry is obtained from observations that are simulated from a standard normal distribution. The true population values of the parameters are given as follows:

$$\Lambda'_0 = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix}$$

$$, \Phi_0 = \begin{bmatrix} 1.0 & 0.3 & 0.3 \\ 0.3 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{bmatrix}$$

$$\Lambda'_0 = \begin{bmatrix} 1.0* & 0.6 & 0.6 & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* \\ 0.0* & 0.0* & 0.0* & 1.0* & 0.6 & 0.6 & 0.0* & 0.0* & 0.0* \\ 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 1.0* & 0.6 & 0.6 \end{bmatrix}$$

$$, \Psi_0 = \text{diag}(0.2, \dots, 0.2)$$

In this paper, parameters with an asterisk are fixed to identify the model. For the measurement equation of the first-level model associated with $t=1,2,3$ (see (3)), no fixed covariates are involved. The true values of parameters in Λ_t and Ψ_t are given by

$$\Lambda'_t = \begin{bmatrix} 1.0* & 0.8 & 0.8 & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* \\ 0.0* & 0.0* & 0.0* & 1.0* & 0.8 & 0.8 & 0.0* & 0.0* & 0.0* \\ 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 0.0* & 1.0* & 0.8 & 0.8 \end{bmatrix}$$

$$, \Psi_t = \text{diag}(0.5, \dots, 0.5)$$

The latent variable in $\omega_{gt} = (\eta_{gt}, \xi_{1gt}, \xi_{2gt})'$ is modeled by the following structural equation:

$$\eta_{gt} = b_{10}d_{1g0} + b_{20}d_{2g0} + b_t d_{gt} + \beta_{t-1} \eta_{g,t-1} + \gamma_{1t} \xi_{1gt} + \gamma_{2t} \xi_{2gt} + \gamma_{3t} \xi_{1gt} \xi_{2gt} + \delta_{gt} \tag{12}$$

where the two components in the fixed covariate vector $\mathbf{d}_{g0} = (d_{1g0}, d_{2g0})'$ are independently generated from the standard bivariate normal distribution, and the fixed covariate d_{gt} ($t=1,2,3$) are generated from Bernoulli distribution with probability of success 0.6. The true values of the parameters involved in (12) are: $b_{10}=1.0, b_{20}=-1.0$; and for all $t=1,2,3$, $b_t=0.7, \beta_{t-1}=0.6, \gamma_{1t}=\gamma_{2t}=0.6, \gamma_{31}=-0.4, \gamma_{32}=-0.6, \gamma_{33}=-0.8, \Psi_{\delta t}=0.5, (\phi_{t11}, \phi_{t12}, \phi_{t22})$ in Φ_{tt} equal to $(1.0, 0.3, 1.0)$,

and covariances of ξ_{igtj} and ξ_{kgtl} are all 0.3. Based on the above specifications of the model and the true parameter values, continuous observations of \mathbf{u}_{gt} as defined by (4) can be obtained. To consider the model with ordered categorical outcomes, the 5th, 6th, and 9th entries in every \mathbf{u}_{gt} were transformed to ordered categorical using the following thresholds (-1.2*, -0.6, 0.6, 1.2*). Moreover, missing data are considered by randomly deleting some entries in \mathbf{u}_{gt} . About 80 per cent of the observations in the data set are fully observed. The total number of parameters in this two-level SEM is 142. In the MCEM algorithm for computing the ML estimates, we generate 100 observations to approximate the conditional expectations at the E-step of each MCEM iteration. Convergence can be monitored by the method of Shi and Copas [40] by using $\gamma_0=3, \delta_1^*=\delta_2^*=0.001$. Convergence is claimed if the stopping rule is satisfied for three consecutive iterations. The algorithm stops at less than 100 MCEM iterations in all 100 replications.

Based on 100 replications, the bias (Bias) and the root mean squares (Rms) between the estimates and the true population values of the parameters were computed as follows:

$$\text{Bias of } \widehat{\theta}(h) = 100^{-1} \sum_{r=1}^{100} \widehat{\theta}_r(h) - \theta_0(h), \quad \text{Rms of } \widehat{\theta}(h) = \left\{ 100^{-1} \sum_{r=1}^{100} [\widehat{\theta}_r(h) - \theta_0(h)]^2 \right\}^{\frac{1}{2}}$$

where $\theta_0(h)$ is the h th element of the true parameter vector, and $\widehat{\theta}_r(h)$ is its ML estimate. Standard error (SE) of the estimates were obtained by the arithmetic mean of SE computed through the Louis formula [13] in 100 replications. Simulation results obtained on the basis of 100 replications are presented in Tables I and II. We observe from these tables that the 'Bias,' 'Rms,' and 'SE' are rather small. These results indicate that the empirical performance of the ML estimates obtained by the MCEM algorithm is satisfactory. To save space, results on the nuisance threshold parameters are not presented.

5. AN APPLICATION OF THE SEM APPROACH: THE LONGITUDINAL STUDY OF COCAINE USE

The data set in this example is obtained from a longitudinal study about cocaine use conducted at the UCLA Center for Advancing Longitudinal Drug Abuse Research. Various measures were collected from patients admitted to the West Los Angeles Veterans Affairs Medical Center in 1988–1989 and who met the DSM III-R criteria for cocaine dependence [44]. These patients were assessed at baseline, one year after treatment, two years after treatment, and 12 years after treatment ($t=t_1, t_2, t_3, t_4$) in 2002–2003. Among these patients located at the 12-year follow-up, some were confirmed to be deceased, some declined to be interviewed, and some were either out of the country or too ill to be interviewed after one year, two years, or at the 12-year follow-up. Hence, there is a considerable amount of missing data. Seven observed variables at each t are involved in the current analysis: (i) cocaine use (CC), an ordered categorical variable with codings 1–5 to denote days of cocaine use per month that are fewer than 2 days, between 2 and 7 days, between 8 and 14 days, between 15 and 25 days, and more than 25 days, respectively; (ii) Beck inventory (BI), an ordered categorical variable with codings 1–5 to denote scores that are less than 3.0, between 3.0 and 8.0, between 9.0 and 20.0, between 21 and 30, and larger than 30; (iii) depression (DEP), an ordered categorical variable based on the Hopkin Symptom Checklist-58 scores, with codings 1–5 to denote scores that are less than 1.1, between 1.1 and 1.4, between 1.4 and 1.8, between 1.8 and 2.5, and larger than 2.50; (iv) number of friends (NF), an ordered categorical variable with codings 1–5 to denote no friend, 1 friend, 2–4 friends,

5–8 friends, more than 9 friends; (v) 'have someone to talk to about problem (TP)'; (vi) 'currently employed (EMP)'; and (vii) 'alcohol dependence (AD) at baseline.' The last three variables are ordered binary variables with {0,1} for {No,Yes}. The sample size is 223, and the frequencies of all variables at different time points are given in Table III.

In this study, we consider observed variables (CC, BI, DEP, NF, TP) and two fixed covariates (EMP, AD) for each individual, which were measured at four time points. The proposed dynamic two-level SEM (see equations (1)–(3) and (5)) is applied to analyze this data set. The second-level model for assessing invariant effects over time is defined by the following two-factor confirmatory factor analysis model:

$$y_g = \mathbf{A}_0 c_{g0} + \Lambda_0 \omega_{g0} + \varepsilon_{g0}, \quad g=1, \dots, 223 \tag{13}$$

where y_g corresponds to (CC, BI, DEP, NF, TP) of the g th individual, c_{g0} is fixed at 1.0 so that \mathbf{A}_0 is a vector of intercepts, $\omega_{g0} = (\omega_{g01}, \omega_{g02})'$, $\Psi_0 = \text{diag}(0.0^*, \psi_{02}, \psi_{03}, \psi_{04}, \psi_{05})$,

$$\Lambda'_0 = \begin{bmatrix} 1.0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & \lambda_{22,0} & \lambda_{32,0} & \lambda_{42,0} & \lambda_{52,0} \end{bmatrix} \quad \text{and} \quad \Phi_0 = \begin{bmatrix} 1.0^* & \\ \phi_{21,0} & 1.0^* \end{bmatrix}$$

In this example, parameters with an asterisk were fixed for specifying an identified model. Note that as y_{g1} is fully defined by CC, $\lambda_{11,0}$ and ψ_{01} are, respectively, fixed at 1.0 and 0.0 according to the usual practice of SEM [2]. In this formulation, ω_{g01} can be interpreted as the invariant portion of cocaine use and ω_{g02} can be interpreted as an invariant general latent factor that is not changed over time. The correlation of ω_{g01} and ω_{g02} is assessed by $\phi_{21,0}$. The first-level model for assessing various dynamic effects that are changed over time is defined by the following nonlinear SEM. The measurement model is defined by

$$v_{gt} = \Lambda_t \omega_{gt} + \varepsilon_{gt}, \quad g=1, \dots, 223, \quad t=t_1, t_2, t_3, t_4 \tag{14}$$

where $\omega_{gt} = (\eta_{gt}, \xi_{1gt}, \xi_{2gt})'$, and

$$\Lambda'_t = \begin{bmatrix} 1.0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 1.0^* & \lambda_{32,t} & 0^* & 0^* \\ 0^* & 0^* & 0^* & 1.0^* & \lambda_{53,t} \end{bmatrix}$$

The nonoverlapping structure of Λ_t is used for achieving a clear interpretation of the latent variables. Based on the meaning of the observed variables and the nonoverlapping structure of Λ_t , latent variables $\eta_{gt}, \xi_{1gt}, \xi_{2gt}$ can be clearly interpreted as 'cocaine use (CC),' 'psychiatric problems,' and 'social support.' In the structural equations of the first-level model, cocaine use was treated as the dependent variable (η_{gt}), and it is regressed on various independent latent variables, together with fixed covariates EMP(d_{gt}) and AD(d_{g0}) that are variant and invariant over time, respectively. More specifically, the structural equations are defined by

$$\begin{aligned} \text{for } t=t_1: & \quad \eta_{g1} = b_0 d_{g0} + b_1 d_{g1} + \gamma_{11} \xi_{1g1} + \gamma_{21} \xi_{2g1} + \delta_{g1} \\ \text{for } t=t_2, t_3, t_4: & \quad \eta_{gt} = b_0 d_{g0} + b_t d_{gt} + \beta_{t-1} \eta_{g,t-1} + \gamma_{1t} \xi_{1gt} + \gamma_{2t} \xi_{2gt} + \delta_{gt} \end{aligned} \tag{15}$$

Clearly, (15) is a special case of (5). So far, this dynamic two-level SEM involves 94 unknown parameters. As the sample size is relatively small ($n=223$), it is not worthwhile to treat the nuisance threshold parameters of the ordered categorical variables as unknown. Hence they

are fixed at $\alpha_{jh} = \Phi^{*-1}(\rho_{jh})$, where Φ^* is the distribution function of $N[0,1]$, and ρ_{jh} are the observed cumulative marginal proportions of the categories with $z_j < h$, see [6, 36].

Let M_1 be the two-level model defined as above. To roughly illustrate model comparison with BIC, we compare it with M_2 , which is a single-level model that is defined by equations (14) and (15) without the second level corresponding to \mathbf{y}_g as given in (13). In the MCEM algorithm for computing the ML estimates under the competing models, we generate $30+5l$ observations to approximate the conditional expectations at the E-step of the l th iteration (see [8]). Based on the stopping rule given in Section 4, the algorithm stops at the 202nd MCEM iteration in M_1 , and $\hat{\theta}^{(202)}$ is taken as the ML estimate in M_1 . Fewer MCEM iterations are required to attain the convergence in the simpler model M_2 . The BIC value for model comparison between M_1 and M_2 is $BIC_{12} = -140.6$. According to the interpretation of BIC [41], this result indicates that the two-level model M_1 is significantly better than the single-level model M_2 . The ML estimates and their standard error estimates (which were computed using 5000 simulated observations) of parameters in the first and second level of the selected model M_1 are, respectively, presented in path diagrams displayed in Figures 1(a) and (b), while those corresponding to Φ are presented in Table IV. Inspired by the model checking technique in regression and the analysis in Lee and Song [8], we use the following estimated residuals to reveal the adequacy of the measurement models and the structural equation in proposed two-level SEM for fitting the

$$\begin{aligned} \widehat{\delta}_{g1} &= \widehat{\eta}_{g1} - \widehat{b}_0 d_{g0} - \widehat{b}_1 d_{g1} - \widehat{\gamma}_{11} \widehat{\xi}_{1g1} \\ \text{data: } \widehat{\varepsilon}_{gt} &= \widehat{\mathbf{u}}_{gt} - \widehat{\mathbf{y}}_g - \widehat{\Lambda}_t \widehat{\omega}_{gt}, \widehat{\varepsilon}_{g0} = \widehat{\mathbf{y}}_g - \widehat{\mathbf{A}}_0 \mathbf{c}_{g0} - \widehat{\Lambda}_0 \widehat{\omega}_{g0}, & - \widehat{\gamma}_{21} \widehat{\xi}_{2g1}, & \text{ and} \\ \widehat{\delta}_{gt} &= \widehat{\eta}_{gt} - \widehat{b}_0 d_{g0} - \widehat{b}_1 d_{gt} - \widehat{\beta}_{t-1} \widehat{\eta}_{g,t-1} \\ & - \widehat{\gamma}_{1t} \widehat{\xi}_{1gt} \\ & - \widehat{\gamma}_{2t} \widehat{\xi}_{2gt}, \end{aligned}$$

for $t = t_2, t_3, t_4$. Plots of the estimated residual in the first level $\widehat{\varepsilon}_{gt}$ versus, $\widehat{\eta}_{gt}$, $\widehat{\xi}_{1gt}$, and $\widehat{\xi}_{2gt}$ and plots of $\widehat{\delta}_{gt}$ versus $\widehat{\xi}_{1gt}$ and $\widehat{\xi}_{2gt}$ at $t = t_1$ are presented in Figure 2. Other estimated residual plots have similar behaviors. These plots lie within two parallel horizontal lines that are centered at zero, and no linear or quadratic trends are detected. This roughly indicates that the proposed measurement models and the structural equation are adequate in fitting the data.

We can conclude from the model comparison result given by BIC_{12} ($= -140.6$) value that it is significantly better to include a second-level model \mathbf{y}_g for incorporating characteristics that are invariant over time to analyze the longitudinal properties of \mathbf{u}_{gt} , see equation (1). It follows from equation (13) and Figure 1(b) that the following findings can be achieved: (i) All the factor loading estimates are significant; this indicates substantial associations of the observed variables with the general latent factor. (ii) The correlation between the general latent factor and cocaine use is not significant. (iii) As expected, the intercept estimates corresponding to the ordered categorical variables are close to zero. The intercept estimate (\widehat{a}_{05}) corresponding to the dichotomous variable y_5 is 0.281. This indicates that $y_5 = 0$ if the underlying latent continuous variable is less than -0.281 ; otherwise $y_5 = 1$. Clearly, the above findings cannot be obtained by a single level model without the multilevel component.

The longitudinal patterns of the more important parameters in the first-level model are displayed in Figures 3(a) and (b). From Figures 1(a) and 3(a) and (b), the following phenomena are observed: (i) As λ_{32} at all time points ($t = t_1, t_2, t_3, t_4$) are close to 1.0, BI and DEP give equal constant loadings over time to the latent variable 'psychiatric problems, ξ_{1gt} .' This indicates that the relationship between 'psychiatric problems' and {BI, DEP} are very stable over time (they both measure depression symptoms). (ii) The fixed covariate AD has a positive effect on 'cocaine use, CC.' (iii) The fixed covariate EMP has a negative effect on 'cocaine use, CC' at all time points (see the line ' b_1 '), which indicates that having a job reduces cocaine

use. Also note that the effect of EMP is more substantial during the long-term follow-up period. (iv) The positive effect of the latent variable 'social support' on cocaine use at baseline changed quickly to negative effects at two years after intake, as well as years after treatment (see the line γ_2). This indicates that social support had a substantial impact on reducing cocaine use during the post-treatment period. (v) Note also that the magnitude of $\widehat{\gamma}_1$ decreases from baseline to one or two years after intake, but rebounds during the long-term follow-up period. This indicates that the impact of 'psychiatric problems' is still strong during a long period of time after treatment, although it may be reduced during treatment or shortly after. (vi) It is interesting to find that the variances ϕ_{11}, ϕ_{22} , and ϕ_{12} are basically constant over time (see Figure 3(b)). (vii) For λ_{53}, γ_1 , and γ_2 , their changes are substantial from baseline to one or two years after intake, but less substantial in the subsequent years (see Figures 3(a) and (b)). (viii) The effects of cocaine use from $t=t_1$ to $t=t_2$, and from $t=t_2$ to $t=t_3$ are -0.074 ($=\widehat{\beta}_1$), and 0.463 ($=\widehat{\beta}_2$), respectively. After a time difference of 10 years, the corresponding effect from $t=t_3$ to $t=t_4$ is reduced to -0.163 ($=\widehat{\beta}_3$) (see Figure 1(a)). Most of these effects are significant. These findings are consistent with prior literature showing a strong stability of cocaine use over time. The present modeling results also demonstrate a positive association of psychiatric problems with cocaine use during two of the three follow-up points. The increasing negative association of social support with cocaine use is discernible at the later follow-up points, which suggests that social support becomes a stronger factor influencing cocaine use when the treatment effects dissipate at the longer follow-up point, and future studies are needed to further this investigation.

6. DISCUSSION

An appealing feature of SEMs is the ability to group the correlated observed variables into latent variables so that instead of dealing with a large number of observed variables, we can utilize the structural equation with a much smaller number of latent variables to assess inter-relationships among them. This feature is particularly advantageous in analyzing multivariate longitudinal data, because the comprehensive model that is simultaneously defined at every time point involves $p \times T$ observed variables, which may be too large for a regression model. Moreover, the interpretations based on latent variables are more clear and precise than those that are based on the original observed variables. The proposed two-level nonlinear SEM with covariates provides a rather general framework for assessing various characteristics of changes with respect to time.

There are a few limitations of the current approach that indicate areas needing future research. First, similar to many other statistical models, the proposed SEM depends on the assumption that the latent variables and the residual errors are normally distributed. As this assumption may be violated for many types of biomedical data, it is necessary to develop robust methods that are less reliant on this assumption [45]. Second, due to the complexities of the proposed two-level nonlinear SEM with ordered categorical variables, which involves observed data with a correlation structure and missing entries, and the fact that a saturated model does not exist in the context of nonlinear SEM, we do not have a classical likelihood ratio test statistic to assess the goodness-of-fit of the proposed model. In a Bayesian approach with the unknown parameters treated as random, goodness-of-fit of a hypothesized model can be assessed by using Bayesian predictive p -values (BPPs) [46], which required integration over the posterior distribution of the unknown parameters in the model. As it is simple to perform numerically using posterior simulation draws of the parameters within the iterations of a Markov chain Monte Carlo algorithm in the Bayesian estimation, the BPPs have been widely applied in Bayesian analyses of some latent variable models in biostatistics (see [47]) as well as many SEMs (see [48] and the references therein). However, there is a conceptual problem in applying BPPs in an ML analysis, because parameters are not considered as random and the posterior

distribution of the unknown parameters is not related to the ML approach. Hence, we propose the BIC for model comparison, and residual plots for model checking (see Figure 2 in the example). Clearly, the development of a more formal statistic for goodness-of-fit of the model in the ML context is an interesting topic for future research. Alternatively, the Bayesian approach can be used. Third, the current ML method cannot handle unordered categorical data. As genotype variables are unordered categorical, it is important to develop methods to handle this kind of data for genetic analysis. Fourth, because the missing data considered here are missing at random, it is necessary to generalize the ML approach in analyzing missing data with a nonignorable missing mechanism.

Acknowledgments

This research was supported by a grant (CUHK 450607) from the Research Grant Council of Hong Kong Special Administration Region and grant P30DA016383 from the National Institute on Drug Abuse, Bethesda, MD. Special thanks to Diane M. Herbeck and David Huang for data preparation and to C. P. Chou for valuable comments in improving the manuscript.

Contract/grant sponsor: Research Grant Council of Hong Kong Special Administration Region; contract/grant number: CUHK 450607

Contract/grant sponsor: National Institute on Drug Abuse; contract/grant number: P30DA016383

APPENDIX A: CONDITIONAL DISTRIBUTIONS AND IMPLEMENTATION OF THE E-STEP

Based on the definition and properties of the current model, it can be shown that

$$p(Y|\Omega_0, \Omega_1, \mathbf{U}, \mathbf{Z}, \alpha, \theta) = \prod_{g=1}^G p(\mathbf{y}_g|\omega_{g0}, \omega_g, \mathbf{u}_g, \theta)$$

where

$$p(\mathbf{y}_g|\omega_{g0}, \omega_g, \mathbf{u}_g, \theta) \stackrel{D}{=} \mathbf{N}\left[\Sigma_v \left\{ \sum_{t=1}^T \Psi_t^{-1} (\mathbf{u}_{gt} - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt}) + \Psi_0^{-1} (\mathbf{A}_0 \mathbf{c}_{g0} + \Lambda_0 \omega_{g0}) \right\}, \Sigma_v\right] \quad \text{with} \quad \Sigma_v = \left(\sum_{t=1}^T \Psi_t^{-1} + \Psi_0^{-1} \right)^{-1} \tag{A1}$$

$$p(\Omega_0|Y, \Omega_1, \mathbf{U}, \mathbf{Z}, \alpha, \theta) = \prod_{g=1}^G p(\omega_{g0}|\mathbf{y}_g, \theta)$$

where

$$p(\omega_{g0}|\mathbf{y}_g, \theta) \stackrel{D}{=} \mathbf{N}[\mathbf{m}_{g0}, \Sigma_{g0}] \tag{A2}$$

in which

$$m_{g0} = \Sigma_{g0} \Lambda_0' \Psi_0^{-1} \mathbf{y}_g \quad \text{and} \quad \Sigma_{g0} = (\Phi_0^{-1} + \Lambda_0' \Psi_0 \Lambda_0)^{-1}$$

$$p(\Omega_1 | \mathbf{Y}, \Omega_0, \mathbf{U}, \mathbf{Z}_{\text{obs}}, \alpha, \theta) = \prod_{g=1}^G p(\omega_g | \mathbf{y}_g, \mathbf{u}_g, \theta)$$

where

$$p(\omega_g | \mathbf{y}_g, \mathbf{u}_g, \theta) \propto \exp \left[-\frac{1}{2} \left\{ \xi_g' \Phi^{-1} \xi_g + \sum_{t=1}^T (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt})' \times \Psi_t^{-1} (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt}) + (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt})' \Psi_{\delta t}^{-1} (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt}) \right\} \right] \tag{A3}$$

$$p(\mathbf{W} | \mathbf{Y}, \Omega_0, \Omega_1, \mathbf{X}, \mathbf{Z}, \alpha, \theta) = p(\mathbf{W} | \mathbf{Y}, \Omega_1, \mathbf{Z}, \alpha, \theta) = \prod_{g=1}^G \prod_{t=1}^T \prod_{j=1}^s p(w_{gtj} | y_{gj}, \omega_{gt}, z_{gtj}, \alpha, \theta)$$

and

$$p(w_{gtj} | y_{gj}, \omega_{gt}, z_{gtj}, \alpha, \theta) \stackrel{D}{=} N \left[y_{gj} + \mathbf{A}'_{tj} \mathbf{c}_{gt} + \Lambda'_{tj} \omega_{gt}, \psi_{tj} \right] I(w_{gtj} \in (\alpha_{j,z_{gtj}}, \alpha_{j,z_{gtj}} + 1]) \tag{A4}$$

where z_{gtj} is j th element of \mathbf{z}_{gt} associated with ordered categorical variable w_{gtj} .

The conditional distributions given in (A1) and (A2) are standard normal distributions, drawing observations from them is straightforward. As (A3) is a nonstandard distribution, the MH algorithm is used to simulate observations from (A3). (A4) involves the univariate truncated normal distribution. Simulating observations from this distribution is done by the inverse distribution method proposed in Devroye [49].

To simulate observations from the target density $p(\boldsymbol{\omega}_g | \cdot)$ in (A3), let $\eta_g = (\eta'_1, \dots, \eta'_T)'$, $\Psi = \text{diag}(\Psi_1, \dots, \Psi_T)$, $\Psi_{\delta} = \text{diag}(\Psi_{\delta 1}, \dots, \Psi_{\delta T})$, λ_{tm} be the m th column of Λ_t ,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \mathbf{A}_T \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \lambda_{11} & 0 & \cdots & 0 & \lambda_{12} & \cdots & \lambda_{1q} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_{21} & \cdots & 0 & 0 & \cdots & 0 & \lambda_{22} & \cdots & \lambda_{2q} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_{T1} & 0 & \cdots & 0 & 0 & \cdots & 0 & \lambda_{T2} & \cdots & \lambda_{Tq} \end{pmatrix}$$

$$\Pi = \begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \Gamma_1 & 0 & \cdots & 0 \\ \mathbf{B}_0 & 0 & \mathbf{B}_2 & \cdots & 0 & \beta_1 & 0 & \cdots & 0 & 0 & 0 & \Gamma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{B}_0 & 0 & 0 & \cdots & \mathbf{B}_r & \beta_1 & \beta_2 & \cdots & \beta_{r-1} & 0 & 0 & 0 & & \Gamma_r \end{pmatrix}$$

and $\Pi_* = \mathbf{I} - \Pi$, where \mathbf{I} is an identity matrix with appropriate dimension. We choose $N[\cdot, \sigma^2 \Sigma_*]$ as the proposal distribution, where $\Sigma_*^{-1} = \Sigma^{-1} + \Lambda' \Psi \Lambda$ and

$$\Sigma_*^{-1} = \begin{bmatrix} \Pi_*' \Psi_\delta^{-1} \Pi_* & -\Pi_*' \Psi_\delta^{-1} \Gamma \Delta \\ -\Delta' \Gamma' \Psi_\delta^{-1} \Pi_* & \Phi^{-1} + \Delta' \Gamma' \Psi_\delta^{-1} \Gamma \Delta \end{bmatrix}$$

where $\Delta = \partial \mathbf{F}_1(\xi_g) / \partial \xi_g | \xi_g = 0$. Let $p(\cdot | \mu_g^*, \sigma^2 \Sigma_*)$ be the density function corresponding to the proposal distribution $N[\mu_g^*, \sigma^2 \Sigma_*]$; the MH algorithm is implemented as follows: At the l th iteration with a current value $\omega_g^{(l)}$, a new candidate ω_g^* is generated from $p(\cdot | \omega_g^{(l)}, \sigma^2 \Sigma_*)$ and accepting this new candidate with probability $\min\{1, p(\omega_g^* | \cdot) / p(\omega_g^{(l)} | \cdot)\}$. The variance σ^2 can be chosen such that the average acceptance rate is approximately 0.25 or more, see [50].

APPENDIX B: DETAILS IN COMPLETING THE M-STEP

Let $\mathbf{A}'_{tj}, \mathbf{A}'_{0j}, \Lambda'_{tj}, \Lambda'_{0j}, \Pi'_{tj}$, and \mathbf{B}_{0j} be the j th rows of $\mathbf{A}_t, \mathbf{A}_0, \Lambda_t, \Lambda_0, \Pi_t$, and \mathbf{B}_0 . The derivatives involved in the system of equations given in $\partial Q / \partial \theta = \mathbf{0}$ are

$$\frac{\partial L_1}{\partial \mathbf{A}_t} = \sum_{g=1}^G \sum_{t=1}^T \Psi_t^{-1} (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt}) \mathbf{c}'_{gt}$$

$$\frac{\partial L_1}{\partial \Lambda_{tj}} = \sum_{g=1}^G \sum_{t=1}^T \psi_{tj}^{-1} (u_{gtj} - y_{gj} - \Lambda_{tj} \omega_{gt}) \omega'_{gt}$$

$$\frac{\partial L_1}{\partial \text{diag}(\Psi_t)} = \frac{1}{2} \text{diag} \left[\sum_{g=1}^G \sum_{t=1}^T \Psi_t^{-1} \left\{ (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt}) \right. \right. \\ \left. \left. \times (\mathbf{u}_{gt} - \mathbf{y}_g - \mathbf{A}_t \mathbf{c}_{gt} - \Lambda_t \omega_{gt})' - \Psi_t \right\} \Psi_t^{-1} \right]$$

$$\frac{\partial L_1}{\partial \Pi_{tj}} = \sum_{g=1}^G \sum_{t=1}^T \psi_{\delta t j}^{-1} (\eta_{gtj} - \mathbf{B}_{0j} \mathbf{d}_{g0} - \Pi_{tj} \mathbf{h}_{gt}) \mathbf{h}'_{gt}$$

$$\frac{\partial L_1}{\partial \mathbf{B}_0} = \sum_{g=1}^G \sum_{t=1}^T \Psi_{\delta t}^{-1} (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt}) \mathbf{d}_{g0}$$

$$\frac{\partial L_1}{\partial \Phi} = \frac{1}{2} \Phi^{-1} \sum_{g=1}^G (\xi_g \xi_g' - \Phi) \Phi^{-1}$$

$$\begin{aligned} & \frac{\partial L_1}{\partial \text{diag}(\Psi_{\delta t})} \\ &= \frac{1}{2} \text{diag} \left[\sum_{g=1}^G \sum_{t=1}^T \Psi_{\delta t}^{-1} \left\{ (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt}) (\eta_{gt} - \mathbf{B}_0 \mathbf{d}_{g0} - \Pi_t \mathbf{h}_{gt})' - \Psi_{\delta t} \right\} \Psi_{\delta t}^{-1} \right] \end{aligned}$$

$$\frac{\partial L_2}{\partial \mathbf{A}_0} = \Psi_0^{-1} \sum_{g=1}^G (\mathbf{y}_g - \mathbf{A}_0 \mathbf{c}_{g0} - \Lambda_0 \omega_{g0}) \mathbf{c}'_{g0}$$

$$\frac{\partial L_2}{\partial \Lambda_{0j}} = \psi_{0j}^{-1} \sum_{g=1}^G (y_{gj} - \mathbf{A}_{0j} \mathbf{c}_{g0} - \Lambda_{0j} \omega_{g0}) \omega'_{g0}$$

$$\begin{aligned} & \frac{\partial L_0}{\partial \text{diag}(\Psi_0)} \\ &= \frac{1}{2} \text{diag} \left[\Psi_0^{-1} \sum_{g=1}^G \left\{ (\mathbf{y}_g - \mathbf{A}_0 \mathbf{c}_{g0} - \Lambda_0 \omega_{g0}) (\mathbf{y}_g - \mathbf{A}_0 \mathbf{c}_{g0} - \Lambda_0 \omega_{g0})' - \Psi_0 \right\} \Psi_0^{-1} \right] \end{aligned}$$

$$\frac{\partial L_2}{\partial \Phi_0} = \frac{1}{2} \Phi_0^{-1} \sum_{g=1}^G (\omega_{g0} \omega'_{g0} - \Phi_0) \Phi_0^{-1}$$

Conditional on other parameters, the solution of each individual equation in $\partial Q / \partial \boldsymbol{\theta} = \mathbf{0}$ can be obtained as follows:

$$\mathbf{A}_t = \left(\sum_{g=1}^G \sum_{t=1}^T \mathbf{c}_{gt} \mathbf{c}'_{gt} \right)^{-1} \sum_{g=1}^G \sum_{t=1}^T E \left[\mathbf{u}_{gt} - \mathbf{y}_g - \Lambda_t \omega_{gt} \mid \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\Lambda'_{ij} = \left(\sum_{g=1}^G \sum_{t=1}^T E \left[\omega_{gt} \omega'_{gt} | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right] \right)^{-1} \\ \times \sum_{g=1}^G \sum_{t=1}^T E \left[\omega_{gt} (u_{gtj} - y_{gj} - \mathbf{A}_{tj} \mathbf{c}_{gt}) | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\psi_{\tau j} = \frac{1}{n} \sum_{g=1}^G \sum_{t=1}^T E \left[(u_{gtj} - y_{gj} - \mathbf{A}_{tj} \mathbf{c}_{gt} - \Lambda_{tj} \omega_{gt})^2 | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\Pi'_{ij} = \left(\sum_{g=1}^G \sum_{t=1}^T E \left[\mathbf{h}_{gt} \mathbf{h}'_{gt} | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right] \right)^{-1} \sum_{g=1}^G \sum_{t=1}^T E \left[\mathbf{h}_{gt} (\eta_{gtj} - \mathbf{B}_{0j} \mathbf{d}_{g0}) | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\mathbf{B}_0 = \left(\sum_{g=1}^G \mathbf{d}_{g0} \mathbf{d}'_{g0} \right)^{-1} \sum_{g=1}^G \sum_{t=1}^T E \left[\eta_{gt} - \Pi_{tj} \mathbf{h}_{gt} | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\Phi = \frac{1}{n} \sum_{g=1}^G E \left[\xi_g \xi'_g | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\psi_{\delta ij} = \frac{1}{n} \sum_{g=1}^G \sum_{t=1}^T E \left[(\eta_{gtj} - \mathbf{B}_{0j} \mathbf{d}_{g0} - \Pi_{gtj} \mathbf{h}_{gt})^2 | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\mathbf{A}_0 = \left(\sum_{g=1}^G \mathbf{c}_{g0} \mathbf{c}'_{g0} \right)^{-1} \sum_{g=1}^G E \left[\mathbf{y}_g - \Lambda_0 \omega_{g0} | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\Lambda'_{0j} = \left(\sum_{g=1}^G E \left[\omega_{g0} \omega'_{g0} | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right] \right)^{-1} \sum_{g=1}^G E \left[\omega_{g0} (y_{gj} - \mathbf{A}_{0j} \mathbf{c}_{g0}) | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\psi_{0j} = \frac{1}{n} \sum_{g=1}^G E \left[(y_{gj} - \mathbf{A}_{0j} \mathbf{c}_{g0} - \Lambda_0 \omega_{g0})^2 | \mathbf{X}, \mathbf{Z}, \alpha, \theta \right]$$

$$\Phi_0 = \frac{1}{n} \sum_{g=1}^G E [\omega_{g0} \omega'_{g0} | \mathbf{X}, \mathbf{Z}, \alpha, \theta]$$

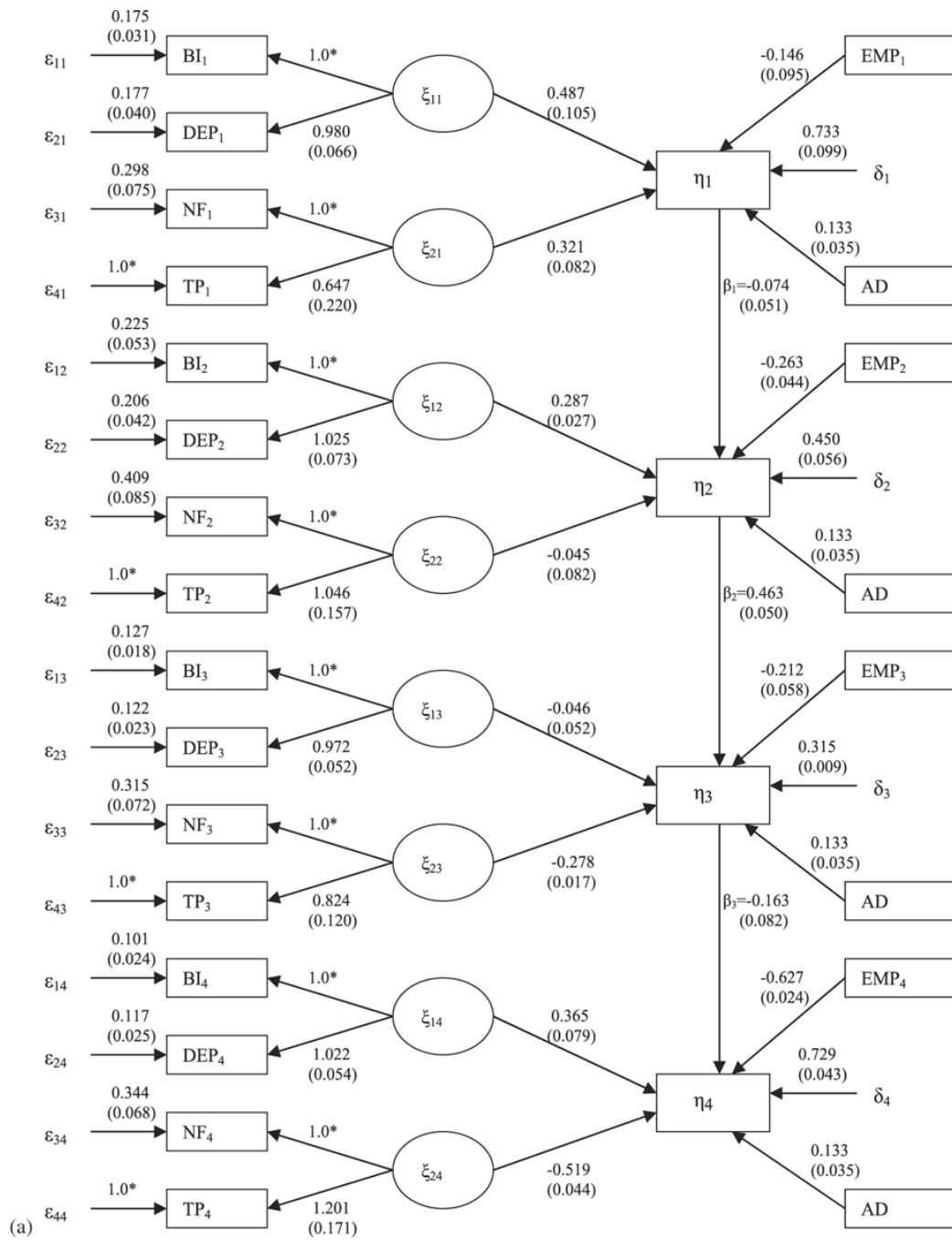
Finally, to update the M-step, the conditional expectations involved in the above solutions are approximated by the corresponding observations simulated by the hybrid algorithm at the E-step.

REFERENCES

1. Bollen, KA. Structural Equation with Latent Variables. Wiley; New York: 1989.
2. Jöreskog, KG.; Sörbom, D. LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Scientific Software International; Hove and London: 1996.
3. Bentler, PM.; Wu, EJC. EQS6 for Windows User Guide. Multivariate Software, Inc.; Enciuo, CA: 2002.
4. Yuan KH, Bentler PM. Mean and covariance structures analysis: the oretical and practical improvements. *Journal of the American Statistical Association* 1997;92:767–774.
5. Dunson DB. Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* 2000;62:355–366.
6. Shi JQ, Lee SY. Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society, Series B* 2000;62:77–87.
7. Lee SY, Shi JQ. Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics* 2001;57:787–794. [PubMed: 11550929]
8. Lee SY, Song XY. Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics* 2004;60:624–636. [PubMed: 15339284]
9. Guo J, Wall M, Amemiya Y. Latent class regression on latent factor. *Biostatistics* 2006;7:145–163. DOI: 10.1093/biostatistics/kxi046. [PubMed: 16079163]
10. Papadopoulos S, Amemiya Y. Correlated samples with fixed and nonnormal latent variables. *Annals of Statistics* 2005;33:2732–2757. DOI: 10.1214/009053605000000552.
11. Sanchez BN, Budtz-Jørgensen E, Ryan LM, Hu H. Structural equation models: a review with applications in environmental epidemiology. *Journal of the American Statistical Association* 2005;100:1443–1454.
12. Wei GCG, Tanner MA. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association* 1990;85:699–704.
13. Louis TA. Finding the observed information matrix when using EM algorithm. *Journal of the Royal Statistical Society, Series B* 1982;44:226–233.
14. Gelman A, Meng XL. Simulating normalizing constant: from importance sampling to bridge sampling to path sampling. *Statistical Science* 1998;13:163–185.
15. Substance Abuse and Mental Health Services Administration. 2002 National Survey on Drug Use & Health Report. SAMHSA, Office of Applied Studies; Rockville, MD: 2003.
16. National Institute of Justice. 2000 Arrestee Drug Abuse Monitoring: Annual Report. NIJ, Office of Justice Programs; Washington, DC: 2003.
17. Substance Abuse and Mental Health Services Administration. Emergency Department Trends from the Drug Abuse Warning Network, Final Estimates 1995–2002. SAMHSA, Office of Applied Studies; Rockville, MD: 2003.
18. Carrol KM, Power ME, Bryant K, Rounsaville BJ. One-year follow-up status of treatment-seeking cocaine abusers. Psychopathology and dependence severity as predictors of outcome. *Journal of Nervous and Mental Disease* 1993;181:71–79. [PubMed: 8426174]
19. Browne RA, Monti PM, Myers MG, Martin RA, Rivinus T, Dubreuil MET, Rohsenow DJ. Depression among cocaine abusers in treatment: relation to cocaine and alcohol use and treatment outcome. *American Journal of Psychiatry* 1998;155:220–225. [PubMed: 9464201]

20. Patkar AA, Thornton CC, Mannelli P, Hill KP, Gottheil E, Vergare MJ, Weinstein SP. Comparison of pretreatment characteristics and treatment outcomes for alcohol, cocaine, and multisubstance-dependent patients. *Journal of Addictive Diseases* 2004;23:93–109.
21. Hayassy BE, Wasserman DA, Hall SM. Social relationships and abstinence from cocaine in an American treatment sample. *Addiction* 1995;90:699–710.
22. Weisner C, Ray GT, Mertens JR, Satre DD, Moore C. Short-term alcohol and drug treatment outcomes predict long-term outcomes. *Drug and Alcohol Dependence* 2003;71:281–294. [PubMed: 12957346]
23. Hser Y. Prediction long-term stable recovery from heroin addiction: findings based on a 33-year follow-up study. *Journal of Addictive Diseases* 2007;26:51–60.
24. Crits-Christoph P, Siqueland L, Blaine J, Frank A, Luborsky L, Onken LS, Muenz LR, Thase ME, Weiss RD, Gastfriend DR, Woody GE, Barber JP, Butler SF, Daley D, Salloum I, Bishop S, Najavits LM, Lis J, Mercer D, Griffin ML, Moras K, Beck AT. Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse Collaborative Cocaine Treatment Study. *Archives of General Psychiatry* 1999;56:493–502. [PubMed: 10359461]
25. Chou CP, Hser YI, Anglin MD. Longitudinal treatment effects among cocaine users: a growth curve modeling approach. *Substance Use and Misuse* 2003;38:1323–1343. [PubMed: 12908814]
26. Hser Y, Stark ME, Paredes A, Huang D, Anglin MD, Rawson R. A 12-year follow-up of a treated cocaine-dependent sample. *Journal of Substance Abuse Treatment* 2006;30:219–226. [PubMed: 16616166]
27. Meredith W, Tisak J. Latent curve analysis. *Psychometrika* 1990;55:107–122.
28. Browne, MW. Structural latent curve models. In: Cuadras, CM.; Rao, CR., editors. *Multivariate Analysis: Future Directions 2*. Elsevier; Amsterdam: 1993. p. 171-179.
29. Dunson DB. Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association* 2003;98:555–563.
30. Daniels MJ, Normand SLT. Longitudinal profiling of health care units based on continuous and discrete patient outcomes. *Biostatistics* 2006;7:1–15. DOI: 10.1093/biostatistics/kxi036. [PubMed: 15917373]
31. Verbeke, G.; Molenberghs, G. *Linear Mixed Models for Longitudinal Data*. Springer; New York: 2000.
32. Diggle, DJ.; Heagerty, P.; Liang, KY.; Zeger, SL. *Analysis of Longitudinal Data*. 2nd edn. Oxford University Press; Oxford, U.K.: 2002.
33. Sammel MD, Ryan LM. Latent variables with fixed effects. *Biometrics* 1996;52:220–243.
34. Rabe-Hesketh S, Skrondal A, Pickless A. Generalized multilevel structural equation modeling. *Psychometrika* 2003;69:167–190.
35. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. Wiley; New York: 1987.
36. Song XY, Lee SY. Bayesian analysis of latent variable models with exponential family outcomes. *Statistics in Medicine* 2007;26:681–693. DOI: 10.1002/sim.2530. [PubMed: 16538704]
37. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984;6:721–741.
38. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of state calculations by fast computing machine. *Journal of Chemical Physics* 1953;21:1087–1091.
39. Hastings WK. Monte Carlo sampling methods using Markov chains and their application. *Biometrika* 1970;57:97–109.
40. Shi JQ, Copas J. Publication bias and meta-analysis for 2x2 tables: an average Markov chain Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* 2002;64:221–236.
41. Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association* 1995;90:773–795.
42. Song XY, Lee SY. Model comparison of generalized linear mixed effect models. *Statistics in Medicine* 2006;25:1685–1698. DOI: 10.1002/sim.2318. [PubMed: 16220521]
43. Meng XL, Wong HW. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* 1996;6:831–860.

44. Kasarabada DN, Anglin MD, Khalsa-Denison E, Paredes A. Differential effects of treatment modality on psychosocial functioning of cocaine-dependent men. *Journal of Clinical Psychology* 1999;55:257–274. [PubMed: 10100826]
45. Lee SY, Lu B, Song XY. Semiparametric Bayesian analysis of structural equation models with fixed covariates. *Statistics in Medicine*. 2008 DOI: 10.1002/sim.3098.
46. Gelman A, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 1996;6:733–807.
47. Carlin JB, Wolfe R, Brown CH, Gelman A. A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. *Biostatistics* 2001;2:397–416. [PubMed: 12933632]
48. Lee, SY. *Structural Equation Modelling: A Bayesian Approach*. Wiley; New York: 2007.
49. Devroye, L. *Non-uniform Random Variate Generation*. Springer; New York: 1985.
50. Gelman, A.; Roberts, GO.; Gilks, WR. Efficient metropolis jumping rules. In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. *Bayesian Statistics*. Vol. vol. 5. Oxford University Press; Oxford: 1995. p. 599-607.



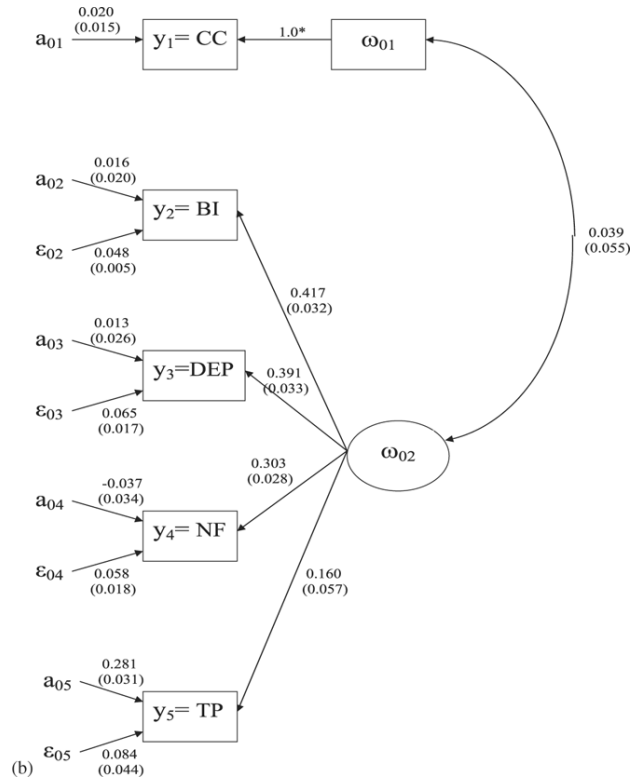


Figure 1.

(a) Path diagram and estimates of some parameters of the first-level model in M_1 , where observed variables are represented by rectangles, latent variables are represented by ellipse, and standard error estimates are in parenthesis. Here, η_t, ξ_{1t} , and ξ_{2t} , respectively, denote 'cocaine use (CC),' 'psychiatric problems,' and 'social support' at $t=1,2,3,4$. (b) Path diagram and estimates of some parameters of the second-level model in M_1 , where observed variables are represented by rectangles, latent variables are represented by ellipse, and standard error estimates are in parenthesis.

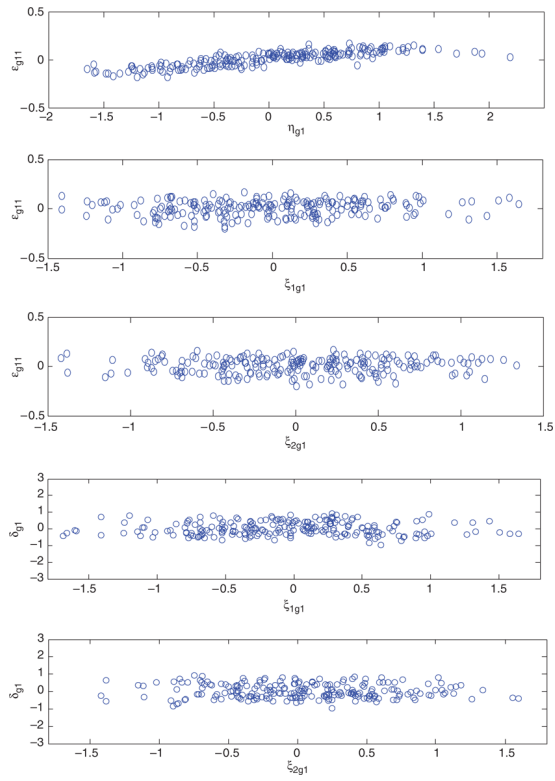


Figure 2.
 Plots of $\hat{\epsilon}_{g11}$ versus $\hat{\eta}_{g1}$, $\hat{\xi}_{1g1}$, and $\hat{\xi}_{2g1}$, and plots of $\hat{\delta}_{g1}$ versus $\hat{\xi}_{1g1}$ and $\hat{\xi}_{2g1}$.

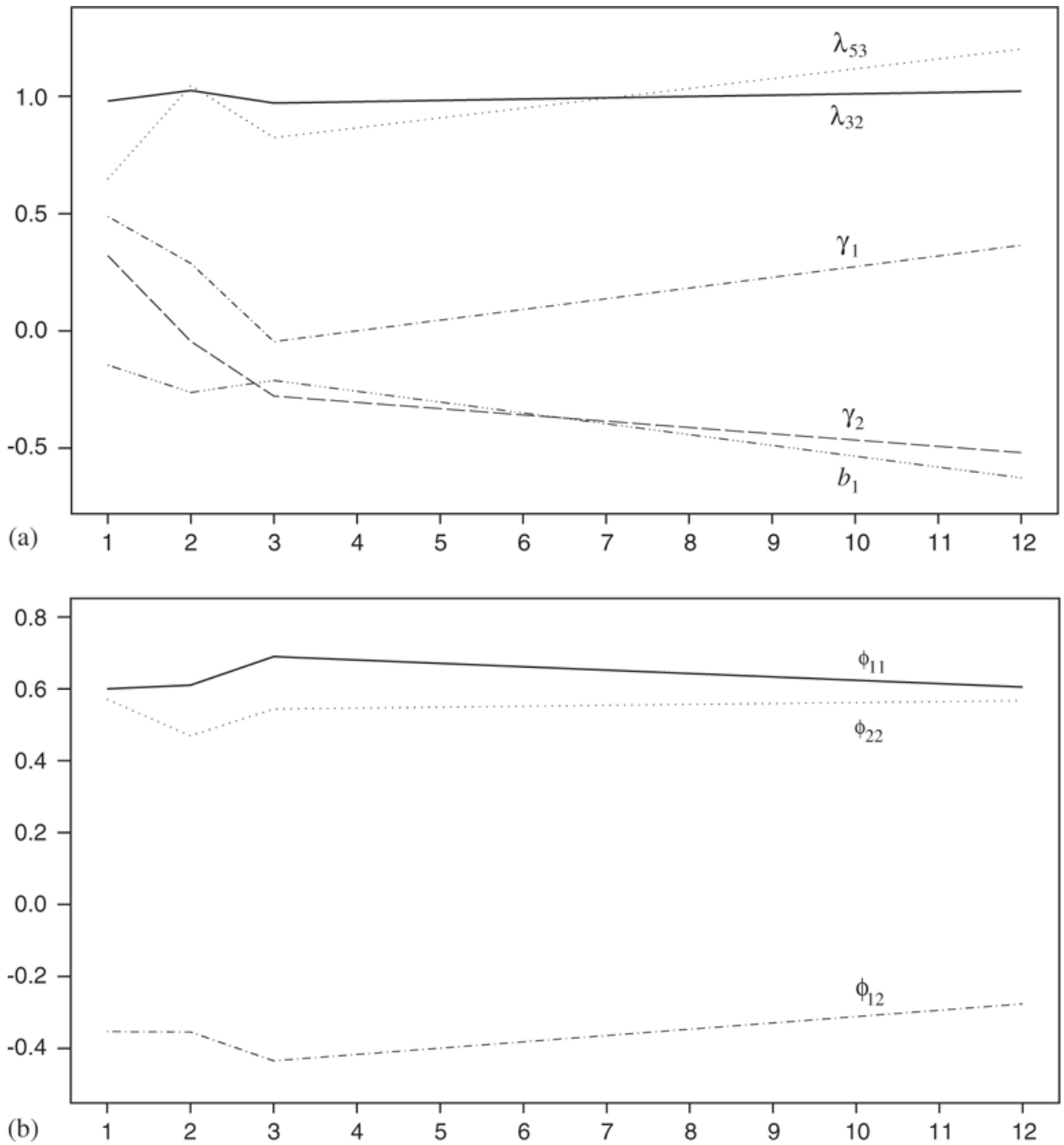


Figure 3. (a) Plots of ML estimates of $\lambda_{32}, \lambda_{53}, b, \gamma_1,$ and γ_2 at $t = t_1, t_2, t_3,$ and t_4 and (b) plots of ML estimates of $\phi_{11}, \phi_{12},$ and ϕ_{22} at $t = t_1, t_2, t_3,$ and t_4 .

Table 1

Means, root mean squares, and standard errors of the parameter estimates in the first level.

First level									
Measurement model					Structural equation				
Par	Bias	Rms	SE	Par	Bias	Rms	SE	Par	SE
$\lambda_{1,21}$	0.004	0.029	0.025	b_{10}	-0.009	0.044	0.019	ϕ_{11}	0.052
$\lambda_{1,31}$	0.004	0.029	0.025	b_{20}	0.017	0.047	0.020	ϕ_{12}	0.003
$\lambda_{1,52}$	-0.004	0.081	0.072	b_1	0.001	0.096	0.079	ϕ_{13}	0.043
$\lambda_{1,62}$	-0.008	0.074	0.057	b_2	0.001	0.079	0.064	ϕ_{14}	0.013
$\lambda_{1,83}$	-0.005	0.068	0.054	b_3	-0.010	0.075	0.070	ϕ_{15}	0.041
$\lambda_{1,93}$	-0.020	0.074	0.063	β_1	0.011	0.042	0.022	ϕ_{16}	0.008
$\lambda_{2,21}$	-0.000	0.018	0.016	β_2	0.004	0.026	0.015	ϕ_{22}	0.023
$\lambda_{2,31}$	-0.001	0.019	0.016	γ_{11}	0.004	0.091	0.053	ϕ_{23}	0.008
$\lambda_{2,52}$	0.001	0.072	0.064	γ_{21}	-0.004	0.085	0.047	ϕ_{24}	0.020
$\lambda_{2,62}$	-0.016	0.068	0.058	γ_{31}	0.034	0.081	0.031	ϕ_{25}	0.010
$\lambda_{2,83}$	0.005	0.066	0.044	γ_{12}	-0.011	0.080	0.047	ϕ_{26}	0.020
$\lambda_{2,93}$	-0.001	0.064	0.065	γ_{22}	-0.015	0.070	0.066	ϕ_{33}	0.066
$\lambda_{3,21}$	-0.001	0.015	0.011	γ_{32}	0.045	0.096	0.047	ϕ_{34}	0.023
$\lambda_{3,31}$	-0.001	0.014	0.014	γ_{13}	-0.022	0.090	0.063	ϕ_{35}	0.049
$\lambda_{3,52}$	-0.002	0.064	0.065	γ_{23}	-0.020	0.078	0.052	ϕ_{36}	0.013
$\lambda_{3,62}$	-0.004	0.071	0.063	γ_{33}	0.054	0.109	0.044	ϕ_{44}	0.037
$\lambda_{3,83}$	-0.007	0.063	0.041					ϕ_{45}	0.017
$\lambda_{3,93}$	0.001	0.071	0.070					ϕ_{46}	0.029
								ϕ_{55}	0.067
								ϕ_{56}	0.014
								ϕ_{66}	0.037
									0.141
									0.082
									0.071
									0.068
									0.061
									0.066
									0.063
									0.104
									0.059
									0.062
									0.064
									0.076
									0.086
									0.061
									0.064
									0.067
									0.091
									0.066
									0.062
									0.082
									0.060
									0.080

Table II

Means, root mean squares, and standard errors of the parameter estimates in the second level.

Par	Bias	Rms	SE	Par	Bias	Rms	SE	Par	Bias	Rms	SE
$\epsilon_{0,11}$	0.008	0.048	0.041	$\epsilon_{0,21}$	-0.039	0.065	0.046	$\lambda_{0,21}$	-0.043	0.065	0.042
$\epsilon_{0,12}$	0.005	0.039	0.032	$\epsilon_{0,22}$	-0.024	0.052	0.035	$\lambda_{0,31}$	-0.037	0.061	0.041
$\epsilon_{0,13}$	0.004	0.039	0.033	$\epsilon_{0,23}$	-0.024	0.044	0.029	$\lambda_{0,52}$	-0.041	0.067	0.049
$\epsilon_{0,14}$	0.008	0.050	0.041	$\epsilon_{0,24}$	-0.022	0.049	0.046	$\lambda_{0,62}$	-0.029	0.055	0.049
$\epsilon_{0,15}$	0.003	0.035	0.074	$\epsilon_{0,25}$	-0.023	0.043	0.038	$\lambda_{0,83}$	-0.034	0.060	0.042
$\epsilon_{0,16}$	0.000	0.042	0.075	$\epsilon_{0,26}$	-0.018	0.044	0.046	$\lambda_{0,93}$	-0.033	0.063	0.047
$\epsilon_{0,17}$	0.002	0.044	0.048	$\epsilon_{0,27}$	-0.020	0.050	0.044	$\phi_{0,11}$	-0.040	0.173	0.099
$\epsilon_{0,18}$	0.004	0.040	0.034	$\epsilon_{0,28}$	-0.015	0.036	0.032	$\phi_{0,12}$	-0.024	0.103	0.073
$\epsilon_{0,19}$	-0.001	0.041	0.062	$\epsilon_{0,29}$	-0.013	0.043	0.040	$\phi_{0,13}$	-0.036	0.105	0.067
								$\phi_{0,22}$	0.008	0.122	0.097
								$\phi_{0,23}$	-0.009	0.077	0.062
								$\phi_{0,33}$	0.015	0.126	0.088

Table III

Frequencies of the ordered categorical variables at different time points in the cocaine use example.

Variables	Categories					Total
	1	2	3	4	5	
$t = 1$						
CC	13	24	38	37	111	223
BI	14	35	68	31	12	160
DEP	15	29	29	57	26	156
NF	27	30	64	27	8	150
TP						200
EMP						223
AD						223
$t = 2$						
CC	89	46	23	22	43	223
BI	30	47	72	24	10	183
DEP	37	46	40	48	18	189
NF	25	27	90	21	18	181
TP						193
EMP						219
AD						223
$t = 3$						
CC	105	30	25	16	47	223
BI	37	44	78	22	11	192
DEP	47	40	47	40	13	187
NF	22	22	94	29	23	190
TP						211
EMP						223
AD						223
$t = 4$						
CC	172	15	14	8	14	223
BI	56	63	72	15	16	222
DEP	63	51	43	46	20	223
NF	21	37	113	26	23	220
TP						220
EMP						223
						198
						139
						223

Variables	Categories					Total		
	1	2	3	4	5		No=0	Yes=1
AD						120	103	223

Table IV

ML estimates and their standard error estimates (in parenthesis) of $\Phi [\Phi_{t_{ij}}]$ in the cocaine use example.

$t = 1$	0.601 (0.054)																		
	-0.354 (0.049)	0.571 (0.056)																	
$t = 2$	0.317 (0.045)	-0.149 (0.029)	0.612 (0.065)																
	-0.283 (0.048)	0.272 (0.045)	-0.355 (0.040)	0.469 (0.051)															
$t = 3$	0.304 (0.053)	-0.426 (0.031)	0.402 (0.061)	-0.324 (0.045)	0.689 (0.048)														
	-0.166 (0.053)	0.374 (0.019)	-0.323 (0.051)	0.271 (0.006)	-0.435 (0.047)	0.544 (0.052)													
$t = 4$	0.100 (0.047)	-0.155 (0.042)	0.093 (0.042)	-0.139 (0.039)	0.204 (0.054)	-0.120 (0.043)	0.605 (0.034)												
	-0.007 (0.039)	0.137 (0.031)	-0.183 (0.056)	0.235 (0.040)	-0.228 (0.051)	0.149 (0.037)	-0.277 (0.040)	0.567 (0.066)											