Chapter 8 Structural Equation Models for Latent Curve Models

- Latent curve models (LCMs) are popular longitudinal methods to analyze individual differences in the patterns of change, which usually involves a random intercept and a random slope with each pair forming a different trajectory over time.
- The random intercept and slope are respectively considered as latent variables representing initial status and rate of change of the outcome variable.

Organization of this chapter:

1. Two motivating examples
2. Several LCMs
   - Basic LCMs
   - LCMs with explanatory latent variables
   - LCMs with longitudinal latent variables
3. Bayesian inference
4. Extensions
   - Nonlinear LCMs
   - Multilevel LCMs
   - Mixture LCMs

Longitudinal study of quality of life of stroke survivors

- Stroke is a major health issue in the older population as it not only affects physical impairment but also leads to disability for activities of daily living (ADL), social non-participation (handicap), and depression. Such changes have a substantial influence on health-related quality of life (HRQOL) of stroke survivors. trajectory.

- Analyses about the dynamic changes of HRQOL of stroke survivors and the associated factors of these changes have received much attention in the field. The existing literature (Ahlsiö *et al.*, 1984; Kwok et al., 2006) found that
  - Even though HRQOL was associated with greater disability, it failed to improve over time even when activities of daily living function increased.
  - Environment and social interaction domains of HRQOL decreased during the first year of stroke.
  - Depression had a more generalized adverse effect on HRQOL than basic functional disabilities.
  - While ADL remained stable, occupation and orientation domains of handicap, and depression deteriorated.

- Research goal:
  - Assess how the changes of daily activities, mental health, and handicap influence the growth of each of the four HRQOL domains across time.

- Data background:
  - Data are obtained from patients with acute stroke within 2 days of admission, and follow-ups at three, six, and twelve months post-stroke.

  - Outcome measures obtained from questionnaires including
    - Modified Barthel Index (MBI) score to assess ADL
    - Geriatric Depression Scale (GDS) score to assess psychological health
    - World Health Organization Quality of Life measure (WHOQOL BREF (HK)) scores to assess health-related quality of life (HRQOL)
    - London Handicap Scale (LHS) to assess handicap

- Patients' ADL were assessed through MBI, which has 10 items on activities of daily living. The total score of 20 indicates full independence in ADL; a higher score represents a higher level of independence.
- Post-stroke depression was measured using GDS with 15 items. Score of 8 or more indicates depression.
- LHS consists of six questions measuring the levels of handicap (scores ranging from 1-6) in six dimensions including mobility, independence, occupation, social integration, orientation, and economic self-sufficiency. Higher score indicates greater handicap. The total score of all six domains was used to indicate overall handicap level.
- The HRQOL dataset includes 24 items in 4 domains: physical health (PHY), psychological health (PSY), social interaction (SO), and environment (EN).
- The stroke survivors are required to respond to the 24 items on a 5-point scale, in which the categories range from "Not at all" to "An extreme amount" (scores ranging from 1-5).

Longitudinal study of cocaine use

Example 2.

- Cocaine use is a major social and health problem in many countries. An interesting topic is to study the dynamic influences of various latent variables on the dynamic change in cocaine use.

- The existing literature (Brown et al., 1998; Hser et al., 2006) has identified several factors that might influence cocaine use:
  - psychiatric problems
  - treatment received
  - family support

Data background:

- The UCLA Center for Advancing Longitudinal Drug Abuse Research collected various measures from patients admitted in 1988-89 to the West Los Angeles Veterans Affairs Medical Center and met the DSM III-R criteria for cocaine dependence.
- These patients were assessed at baseline, one year after treatment, two years after treatment, and 12 years ($t = 1,\ 2,\ 3,\ 4$) after treatment in 2002.
- Measures at each time point include
  - days of cocaine use per month (CC),
  - times per month in formal treatment (TR),
  - Beck Inventory (BI),
  - depression (DEP),
  - anxiety (AN),
  - family support (FS),

- We first consider an LCM based on cocaine use measured at $t = 1, 2, 3, 4$ with a latent variable η_1 on initial status (intercept) of cocaine use, and a latent variable η_2 on rate of change (slope) of cocaine use.
- We further consider LCMs based on treatment (TR), family support (FS), and psychiatric problem (PP) measured at $t = 1, 2, 3, 4$ with latent variables $\{\xi_{11}, \xi_{12}, \xi_{13}\}$ and $\{\xi_{21}, \xi_{22}, \xi_{23}\}$ on the initial status and rate of change of TR, FS, and DEP, respectively.
- The effects of $\{\xi_{11}, \xi_{21}\}$, $\{\xi_{12}, \xi_{22}\}$, and $\{\xi_{13}, \xi_{23}\}$ on $\{\eta_1, \eta_2\}$ are assessed through a structural equation.
- Psychiatric problem (PP) is assessed by three observed variables, including beck inventory (BI), depression (DEP), and anxiety (AN).
- At each time point we first use a measurement equation to group BI, DEP, and AN into the latent variable 'PP', then construct an LCM based on PP at $t = 1, 2, 3, 4$ to obtain the initial status of PP (ξ_{13}) and rate of change of PP (ξ_{23}).

- A latent variable obtained from several observed variables is called a "first-order latent variable" (e.g., η_1, η_2, ξ_{11}, ξ_{21}, ξ_{12}, ξ_{22}); and a latent variable obtained from several latent variables (e.g., ξ_{13}, ξ_{23}) is called a "second-order latent variable" (Jöreskog, 1970).

- It is desirable to establish a comprehensive LCM, which
  - accommodates the effects of first- and second-order latent variables on the growth factors of the outcome variables
  - considers the interaction of latent variables to assess the joint effects of dynamic latent variables
  - includes mixed continuous and ordered categorical data
  - accounts for missing data.

The basic LCM (Meredith and Tisak, 1990; Bollen and Curran, 2006) can be viewed as the following common factor analysis model:

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}_y, \tag{1}$$

where

- $\mathbf{y}$ is a $T \times 1$ vector of repeated measures
- $\boldsymbol{\Lambda}$ is a $T \times m$ parameter matrix of sequential fixed values of the growth curve records. The pattern of $\boldsymbol{\Lambda}$ can represent a particular aspect of change in y across the T occasions.
- $\boldsymbol{\eta}$ is an $m \times 1$ outcome latent growth factor, which contains scores on the m factors for a given individual
- $\boldsymbol{\epsilon}_y$ is a $T \times 1$ vector of residual errors
- $\boldsymbol{\epsilon}_y \sim N[\mathbf{0}, \boldsymbol{\Psi}_y]$, where $\boldsymbol{\Psi}_y$ is a diagonal matrix with diagonal elements ψ_{yt}, for $t = 1, \cdots, T$.

When $m = 2$, Equation (1) is expressed in the following matrix form:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_T \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{y1} \\ \epsilon_{y2} \\ \vdots \\ \epsilon_{yT} \end{pmatrix}. \tag{2}$$

- The first column in $\mathbf{\Lambda}$ can be used to define an intercept factor, which represents an initial status of change in y.
- The second column in $\mathbf{\Lambda}$ represents the known times of measurement (t_1 through t_T) and constraints (the values of t should reflect the spacing between measurement occasions).
- The latent growth factor $\boldsymbol{\eta}$ contains the initial status η_1 (intercept) and rate of change η_2 (slope).

- The two-factor linear latent curve model is specified in (2), so that the intercept factor (fixed at a constant value of 1) serves as the starting point (initial status) for any change (growth) across time and the slope factor captures the rate of change of the trajectory over time.

- The scaling of the slope can be specified by using either fixed value restrictions (e.g. $0, 1, \cdots, T-1$) representing a straight line growth, or unspecified value restrictions (where $t_1 = 0$ and $t_2 = 1$ are fixed for model identification, and the remaining $t_3, \cdots, t_T$ are freely estimated) allowing estimation of an optimal pattern of change over measurement occasions (Meredith and Tisak, 1990).

The equation for the latent growth factor $\boldsymbol{\eta}$ can be simply written as

$$\boldsymbol{\eta} = \boldsymbol{\beta} + \boldsymbol{\Pi}\boldsymbol{\eta} + \boldsymbol{\delta}, \tag{3}$$

where

- $\boldsymbol{\beta}$ is an $m \times 1$ vector of the population average of the latent individual growth factors
- $\boldsymbol{\Pi}$ is an $m \times m$ matrix of coefficients expressing the structural relations between the $\boldsymbol{\eta}$ variables
- $\boldsymbol{\delta}$ is an $m \times 1$ vector of random residuals
- $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Psi}_\eta)$, and $\boldsymbol{\Psi}_\eta$ is a diagonal covariance matrix.

To illustrate the model framework, we use the example in the longitudinal study of cocaine use across four measurement occasions. Let $\mathbf{y} = (y_1, y_2, y_3, y_4)^T$, and $\mathbf{\Pi} = \mathbf{0}$, then Equations (2) and (3) can be expressed in the following forms to represent the model in Figure 8.1:

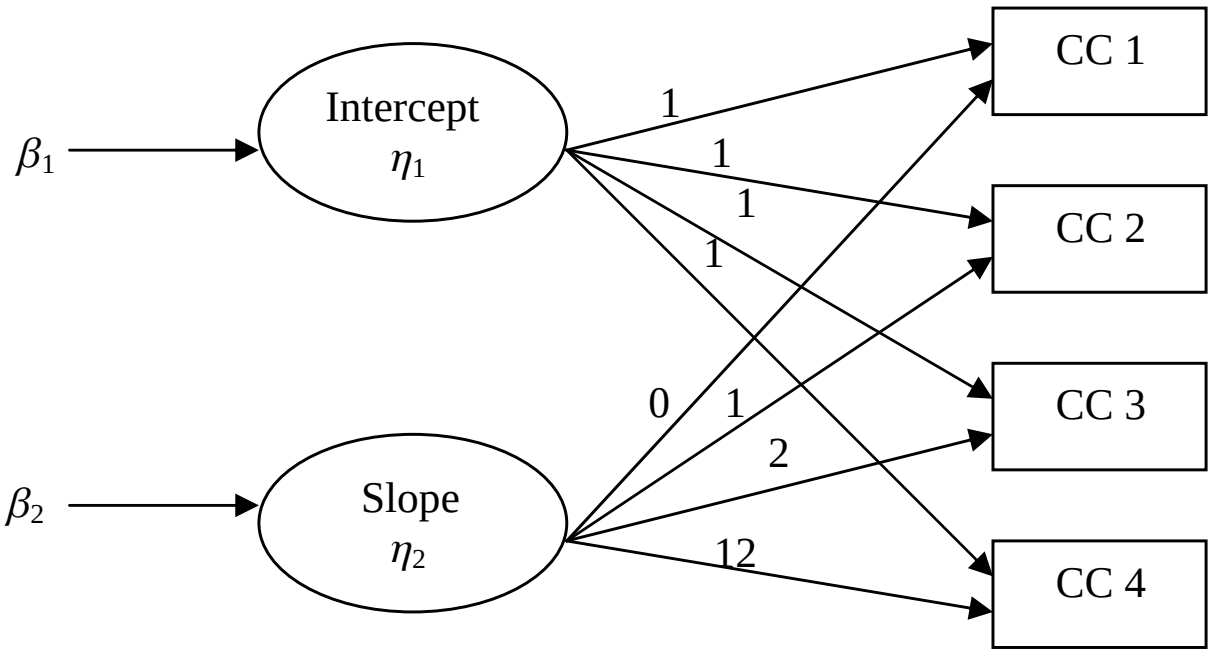
r.v.s

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 12 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{y1} \\ \epsilon_{y2} \\ \epsilon_{y3} \\ \epsilon_{y4} \end{pmatrix}, \tag{4}$$

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}. \tag{5}$$

The four measurement occasions are 1989 (baseline), 1990 (one year after treatment), 1991 (two years after treatment), and 2002 (12 years after treatment). Thus, the fixed times of measurement t_1, t_2, t_3, and t_4 are set to be 0, 1, 2, and 12 to reflect the spacing between measurement occasions.

The path diagram of the aforementioned basic LCM is depicted as follows:



The previous model can be extended to include the following explanatory latent growth factors:

the previous model is a simple one for toy.

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_r \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Lambda} \end{pmatrix} \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \\ \vdots \\ \boldsymbol{\xi}_r \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_{x1} \\ \boldsymbol{\epsilon}_{x2} \\ \vdots \\ \boldsymbol{\epsilon}_{xr} \end{pmatrix}, \tag{6}$$

r

$\mathbf{x}_k \in \mathbb{R}^{T \times 1}$

where

- $\mathbf{0}$ is a $T \times m$ matrix with elements 0
- for $k = 1, \cdots, r$, $\boldsymbol{\xi}_k$ is the kth latent growth factor that includes growth parameters (e.g., intercept, slope), which are the first-order explanatory latent variables
- $\mathbf{x}_k$ and $\boldsymbol{\epsilon}_{xk}$ are similarly defined as $\mathbf{y}$ and $\boldsymbol{\epsilon}_y$, and $\mathbf{\Lambda}$ is similarly defined as in Equation (1)
- $\boldsymbol{\epsilon}_{xk} \sim N[\mathbf{0}, \mathbf{\Psi}_{xk}]$, where $\mathbf{\Psi}_{xk}$ is a diagonal matrix with diagonal elements $\psi_{x1k}, \cdots, \psi_{xTk}$.

In the cocaine use data, the outcome variable is 'cocaine use (CC)', and η_1 and η_2 are the latent growth parameters, where η_1 (intercept) specifies the initial status of individual growth, whereas η_2 (slope) specifies the rate of change over the four measurement time points.

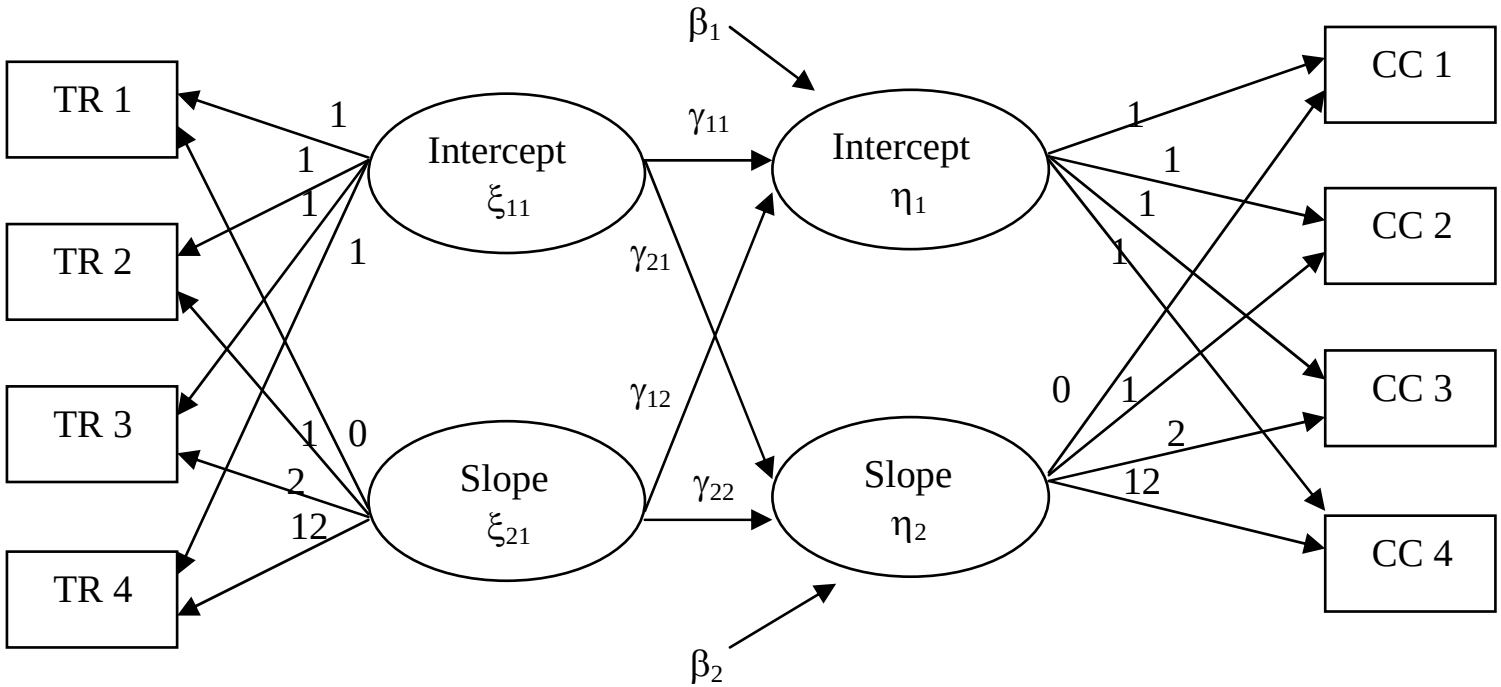
Let $\mathbf{x}_1 = (x_{11}, x_{21}, x_{31}, x_{41})^T$ be a vector of treatment measured at four time occasions, and $\boldsymbol{\xi}_1 = (\xi_{11}, \xi_{21})^T$ such that

$$\begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \\ x_{41} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 12 \end{pmatrix} \begin{pmatrix} \xi_{11} \\ \xi_{21} \end{pmatrix} + \begin{pmatrix} \epsilon_{x11} \\ \epsilon_{x21} \\ \epsilon_{x31} \\ \epsilon_{x41} \end{pmatrix}. \tag{7}$$

The growth factor in relation to treatment, $\boldsymbol{\xi}_1$, can be regarded as an explanatory latent growth factor with intercept (ξ_{11}) and slope (ξ_{21}). The LCM with this explanatory latent growth factor can be expressed by Equations (4) and (7), and the following equation:

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} \xi_{11} \\ \xi_{21} \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}. \tag{8}$$

The path diagram of the above LCM is depicted in Figure 8.2.



To incorporate longitudinal latent variables, we further consider the following measurement model (Song, Lee and Hser, 2009):

$$\mathbf{u}_t = \mathbf{\Lambda}_{ut}\boldsymbol{\omega}_{ut} + \boldsymbol{\epsilon}_{ut}, \quad t = 1, \cdots, T, \tag{9}$$

where

- $\mathbf{u}_t$ is a $p \times 1$ vector of measurements,
- $\mathbf{\Lambda}_{ut}$ is a $p \times q$ loading matrix,
- $\boldsymbol{\omega}_{ut}$ is a $q \times 1$ vector of latent variables,
- $\boldsymbol{\epsilon}_{ut}$ is a $p \times 1$ vector of unique variances in $\mathbf{u}_t$,
- $\boldsymbol{\epsilon}_{ut} \sim N[\mathbf{0}, \mathbf{\Psi}_{ut}]$, where $\mathbf{\Psi}_{ut}$ is a diagonal matrix with diagonal elements $\psi_{ut1}, \cdots, \psi_{utp}$.

Equation (9) can be expressed in the following form:

At t's timepoint.

$$\begin{pmatrix} u_{t1} \\ u_{t2} \\ \vdots \\ u_{tp} \end{pmatrix} = \begin{pmatrix} \lambda_{ut,11} & \lambda_{ut,12} & \cdots & \lambda_{ut,1q} \\ \lambda_{ut,21} & \lambda_{ut,22} & \cdots & \lambda_{ut,2q} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{ut,p1} & \lambda_{ut,p2} & \cdots & \lambda_{ut,pq} \end{pmatrix} \begin{pmatrix} \omega_{ut1} \\ \omega_{ut2} \\ \vdots \\ \omega_{utq} \end{pmatrix} + \begin{pmatrix} \epsilon_{ut1} \\ \epsilon_{ut2} \\ \vdots \\ \epsilon_{utp} \end{pmatrix}, \quad (10)$$

FA at t timepoint.

where some $\lambda_{ut,ij}$ are fixed at 0 or 1 for model identification.

As the latent variables ω_{utk}, $k = 1, \cdots, q$ defined in Equation (9) or (10) are related to observed variables measured at multiple time points $t_1, \cdots, t_T$, they are regarded as longitudinal latent variables.

Let $\boldsymbol{\omega}_k = (\omega_{u1k}, \cdots, \omega_{uTk})^T$, and $\boldsymbol{\epsilon}_{\omega k} = (\epsilon_{\omega 1k}, \cdots, \epsilon_{\omega Tk})^T$, we can further use the following LCM to investigate the patterns of change in these longitudinal latent variables:

previously, r latent variables have been defined as latent factors as treatment variable $x_{1:r}$

$$\boldsymbol{\omega}_k = \boldsymbol{\Lambda}\boldsymbol{\xi}_{r+k} + \boldsymbol{\epsilon}_{\omega k}, \quad k = 1, \cdots, q, \tag{11}$$

where

- $\boldsymbol{\xi}_{r+k}$ is an $m \times 1$ latent growth factor that includes growth parameters,
- r is specified in Equation (6),
- $\boldsymbol{\Lambda}$ is similarly defined as in Equation (1),
- $\boldsymbol{\epsilon}_{\omega k} \sim N[\mathbf{0}, \boldsymbol{\Psi}_{\omega k}]$, where $\boldsymbol{\Psi}_{\omega k}$ is a diagonal matrix with diagonal elements $\psi_{\omega tk}$, $t = 1, \cdots, T$, $k = 1, \cdots, q$.

If $m = 2$, then $\boldsymbol{\xi}_{r+k} = (\xi_{1,r+k}, \xi_{2,r+k})^T$, and Equation (11) has the form:

$$\begin{pmatrix} \omega_{u1k} \\ \omega_{u2k} \\ \vdots \\ \omega_{uTk} \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_T \end{pmatrix} \begin{pmatrix} \xi_{1,r+k} \\ \xi_{2,r+k} \end{pmatrix} + \begin{pmatrix} \epsilon_{\omega 1k} \\ \epsilon_{\omega 2k} \\ \vdots \\ \epsilon_{\omega Tk} \end{pmatrix}, \quad k = 1, \cdots, q, \qquad (12)$$

where $\xi_{1,r+k}$ and $\xi_{2,r+k}$ indicate the intercept and slope in latent growth factor $\boldsymbol{\xi}_{r+k}$.

In the cocaine use example, 'psychiatric problem' is a longitudinal latent variable, which is formed by three observed variables, BI, DEP, and AN, at multiple time points. The measurement equation is given as follows:

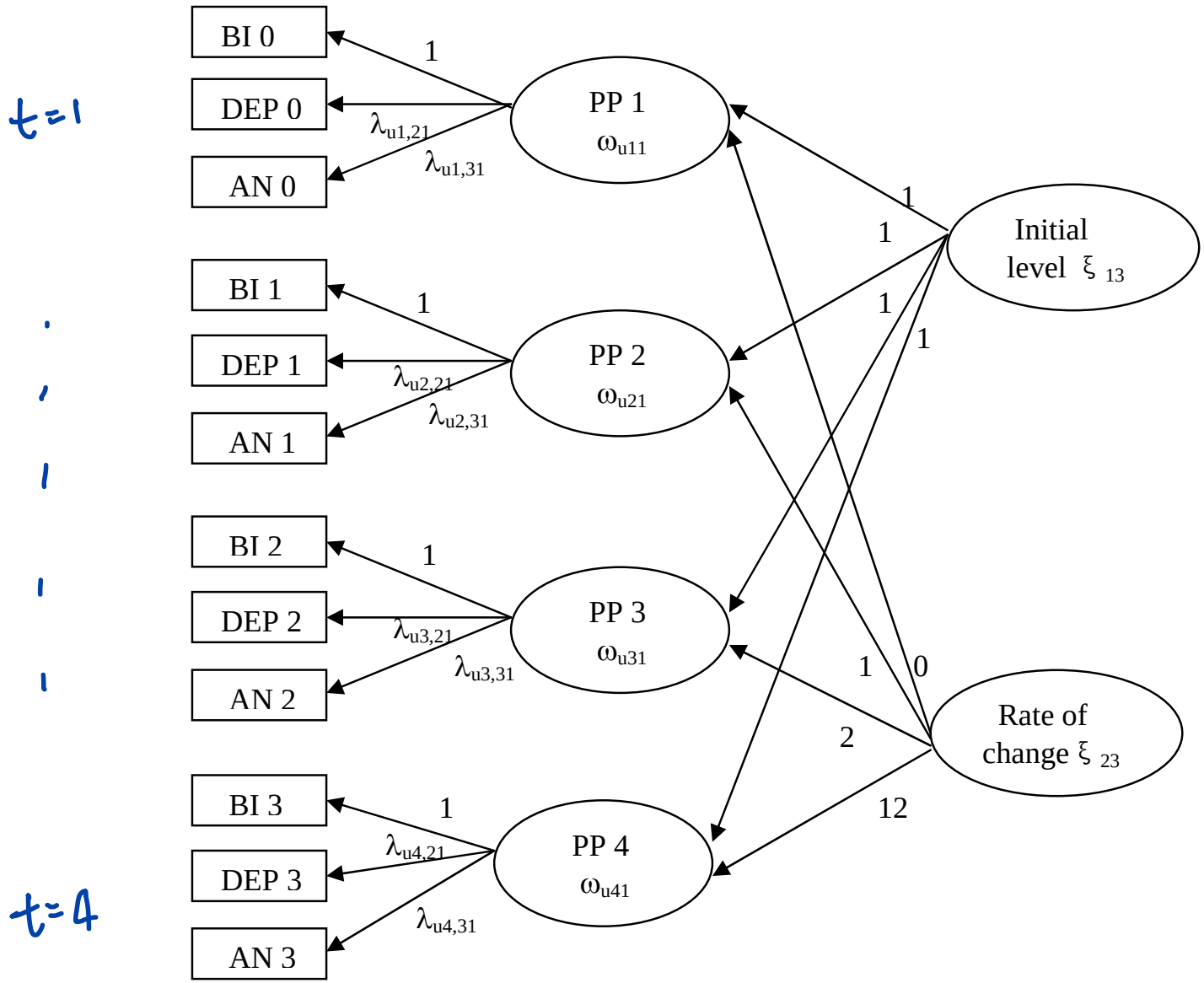
$$\begin{pmatrix} u_{t1} \\ u_{t2} \\ u_{t3} \end{pmatrix} = \begin{pmatrix} 1 \\ \lambda_{ut,21} \\ \lambda_{ut,31} \end{pmatrix} \omega_{ut1} + \begin{pmatrix} \epsilon_{ut1} \\ \epsilon_{ut2} \\ \epsilon_{ut3} \end{pmatrix}, \quad t = 1, \cdots, 4, \qquad (13)$$

where 1's in $\boldsymbol{\Lambda}_{ut}$ are fixed to identify the factor analysis model.

The LCM (see Equation (12)) that describes the pattern of change of the longitudinal latent variable ω_{ut1} is defined as:

$$\begin{pmatrix} \omega_{u11} \\ \omega_{u21} \\ \omega_{u31} \\ \omega_{u41} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 12 \end{pmatrix} \begin{pmatrix} \xi_{13} \\ \xi_{23} \end{pmatrix} + \begin{pmatrix} \epsilon_{\omega 11} \\ \epsilon_{\omega 21} \\ \epsilon_{\omega 31} \\ \epsilon_{\omega 41} \end{pmatrix}, \tag{14}$$

where ξ_{13} and ξ_{23} can be regarded as the second-order latent variables. The path diagram in relation to Equations (13) and (14) is depicted in the following Figure 8.3.


BI 0
DEP 0
AN 0
BI 1
DEP 1
AN 1
BI 2
DEP 2
AN 2
BI 3
DEP 3
AN 3
PP 1
ωu11
PP 2
ωu21
PP 3
ωu31
PP 4
ωu41
Initial level ξ 13
Rate of change ξ 23
λu1,21
λu1,31
λu2,21
λu2,31
λu3,21
λu3,31
λu4,21
λu4,31
1
0
2
12
t=1
t=4

Let $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^T, \cdots, \boldsymbol{\xi}_r^T, \boldsymbol{\xi}_{r+1}^T, \cdots, \boldsymbol{\xi}_{r+q}^T)^T$ be the vector of explanatory latent growth factors that are either first- or second-order latent variables.

To assess the linear and nonlinear effects of the explanatory latent growth factors in $\boldsymbol{\xi}$ on the latent growth factor $\boldsymbol{\eta}$, we further propose the following nonlinear structural equation:

$$\boldsymbol{\eta} = \boldsymbol{\beta} + \boldsymbol{\Pi}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{F}(\boldsymbol{\xi}) + \boldsymbol{\delta}, \tag{15}$$

where

- $\boldsymbol{\Gamma}$ is a matrix of unknown regression coefficients,
- $\mathbf{F}(\boldsymbol{\xi}) = (f_1(\boldsymbol{\xi}), \cdots, f_h(\boldsymbol{\xi}))^T$ is a vector-valued function containing nonzero differentiable functions $f_1, \cdots, f_h$, and $h \geq r + q$,
- $\boldsymbol{\xi} \sim N[\mathbf{0}, \boldsymbol{\Phi}]$, and $\boldsymbol{\Phi}$ contains variances and covariances of explanatory latent growth factors
- $\boldsymbol{\delta} \sim N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$, in which $\boldsymbol{\Psi}_\delta$ is a diagonal matrix with diagonal elements $\psi_{\delta k}$, $k = 1, \cdots, m$, and $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ are independent

Let $\mathbf{\Lambda}_\delta = (\mathbf{\Pi}, \mathbf{\Gamma})$, $\boldsymbol{\zeta} = (\boldsymbol{\eta}^T, \boldsymbol{\xi}^T)^T$, and $\mathbf{G}(\boldsymbol{\zeta}) = (\boldsymbol{\eta}^T, \mathbf{F}(\boldsymbol{\xi})^T)^T$, then Equation (15) can be rewritten as

$$\boldsymbol{\eta} = \boldsymbol{\beta} + \mathbf{\Lambda}_\delta \mathbf{G}(\boldsymbol{\zeta}) + \boldsymbol{\delta}. \tag{16}$$

In the analysis of cocaine use data, we are interested in the longitudinal effects of treatment (TR), family support (FS), and psychiatric problem (PP) on cocaine use (CC), where 'psychiatric problem (PP)' is a latent variable.

We aim to study how the dynamic changes of the latent growth factors $\boldsymbol{\xi}_1$ (TR), $\boldsymbol{\xi}_2$ (FS), and $\boldsymbol{\xi}_3$ (PP) influence the latent growth factor $\boldsymbol{\eta}$ (CC). Here, $m = 2$, $r = 2$, $q = 1$, and each $\boldsymbol{\xi}_k = (\xi_{1k}, \xi_{2k})^T$ $(k = 1,\ 2,\ 3)$ includes the intercept and slope that represent the initial status and rate of changes in TR, FS, and PP, respectively.

An interesting model is described as follows:

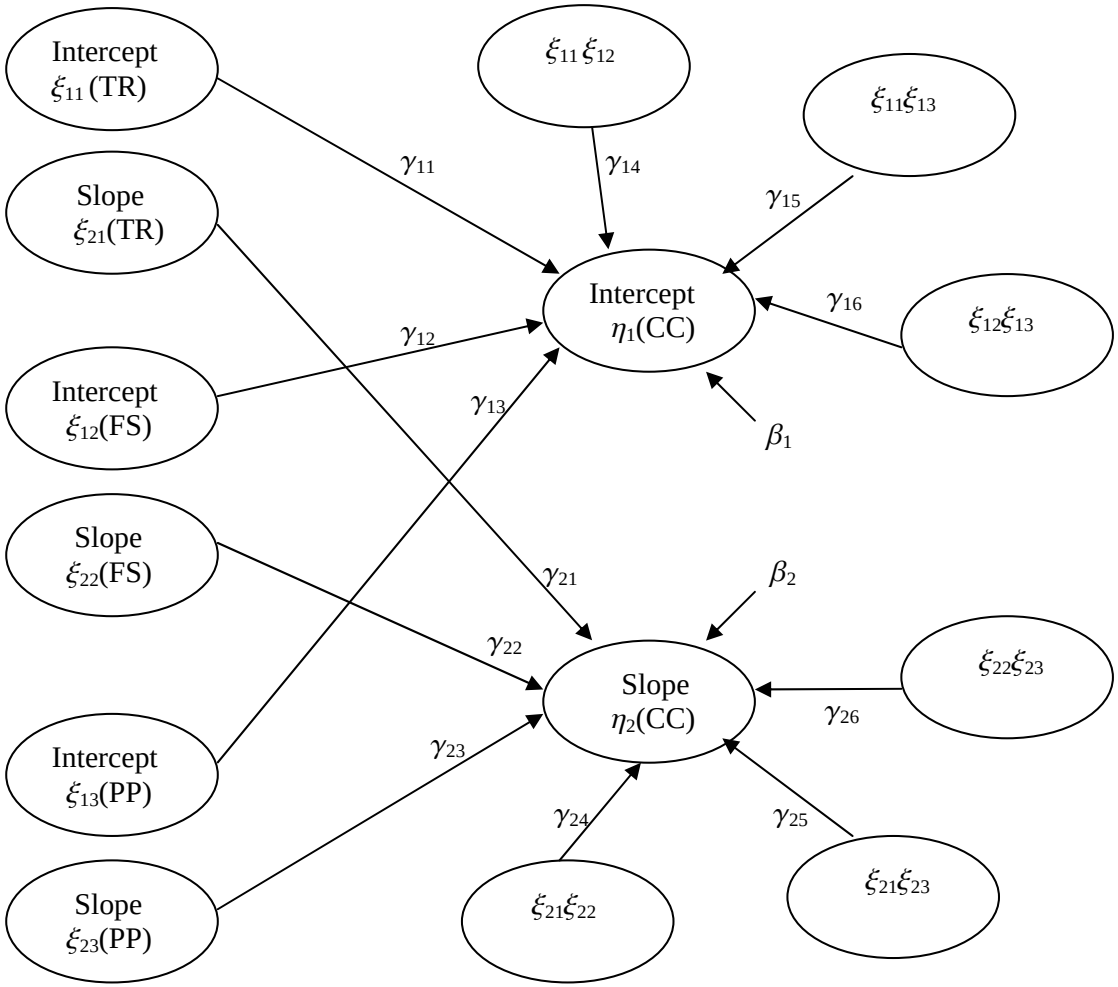
$$\eta_1 = \beta_1 + \gamma_{11}\xi_{11} + \gamma_{12}\xi_{12} + \gamma_{13}\xi_{13} + \gamma_{14}\xi_{11}\xi_{12} + \gamma_{15}\xi_{11}\xi_{13} + \gamma_{16}\xi_{12}\xi_{13} + \delta_1, \tag{17}$$

$$\eta_2 = \beta_2 + \gamma_{21}\xi_{21} + \gamma_{22}\xi_{22} + \gamma_{23}\xi_{23} + \gamma_{24}\xi_{21}\xi_{22} + \gamma_{25}\xi_{21}\xi_{23} + \gamma_{26}\xi_{22}\xi_{23} + \delta_2, \tag{18}$$

where

- γ_{11}, γ_{12}, and γ_{13} represent the effects of initial status in treatment, family support, and psychiatric problem on the initial status in cocaine use,
- γ_{21}, γ_{22}, and γ_{23} represent the effects of rate of change in treatment, family support, and psychiatric problem on the rate of change in cocaine use,
- other coefficients represent various interactive effects of explanatory growth factors on the outcome growth factor.

The path diagram of the above model is depicted in Figure 8.4.



For example, Equation (18)

$$\eta_2 = \beta_2 + \gamma_{21}\xi_{21} + \gamma_{22}\xi_{22} + \gamma_{23}\xi_{23} + \gamma_{24}\xi_{21}\xi_{22} + \gamma_{25}\xi_{21}\xi_{23} + \gamma_{26}\xi_{22}\xi_{23} + \delta_2$$

can be used to examine the impact of joint changes in TR, FS, and PP influences the change in variable CC.

Hence, the significance of parameters γ_{24}, γ_{25}, and γ_{26} essentially answers the question: Does the simultaneous change (interaction effect) in explanatory latent growth factors in treatment, psychiatric problem, or family support influence the rate of change in the outcome growth factor in cocaine use?

So far we have focused our attention on continuous repeated measures. However, in the cocaine use study, some of the observed variables y, x, and u are ordered categorical variables. These ordered categorical variables will be treated similarly as before.

Let z be the underlying continuous unobservable variable that corresponds to an ordered categorical variable x. Then, for $k = 1, \cdots, r$, $t = 1, \cdots, T$:

for every time point t.

$$x_{tk} = j, \quad \text{if} \quad \alpha_{tk,j} \leq z_{tk} < \alpha_{tk,j+1}, \quad j = 0, \cdots, h_k, \tag{19}$$

where $\{-\infty = \alpha_{tk,0} < \alpha_{tk,1} < \cdots < \alpha_{tk,h_k} < \alpha_{tk,h_k+1} = \infty\}$ is the set of threshold parameters that defines the $h_k + 1$ categories. For identification, we follow the suggestion given in Bollen and Curran (2006) to set the first threshold $\alpha_{tk,1}$ to 0 and ψ_{xtk} to 1 for $k = 1, \cdots, r$, $t = 1, 2, \cdots, T$.

To utilize the Bayesian approach to analyze the above LCMs, let $\boldsymbol{\theta}$ be the vector that contains the unknown free parameters of the model, and $\boldsymbol{\alpha}$ be the vector that contains the unknown thresholds. The parameter vector $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ is considered as random with a prior distribution $p(\boldsymbol{\theta}, \boldsymbol{\alpha})$.

Let $\mathbf{D} = \{\mathbf{Y}, \mathbf{X}, \mathbf{U}\}$, where $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{U}$ include all three kinds of observed variables defined in (2), (6), and (10). $\mathbf{Z}$ includes all the underlying continuous variables that correspond to ordered categorical variables in $\mathbf{D}$.

The Bayesian inference is based on $p(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathbf{D})$. For the proposed LCM, directly handling $p(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathbf{D})$ is difficult due to the existence of first- and second-order latent variables, unobservable continuous variables, and the nonlinearity of the structural equation. We adopt the commonly used strategy that involves two key steps to solve the difficulty.

The first step is to augment the observed data with all the latent variables to form a complete data set. Let $\boldsymbol{\Xi}$ include all latent growth factors ($\boldsymbol{\eta}$ and $\boldsymbol{\xi}$), $\boldsymbol{\Omega}$ include all longitudinal latent variables. The complete data set is $\{\mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{\Omega}, \mathbf{D}\}$. Treating $\mathbf{Z}$, $\boldsymbol{\Xi}$, and $\boldsymbol{\Omega}$ as random unknown quantities, the joint posterior density is $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{\Omega}|\mathbf{D})$.

The second step is to draw a sufficiently large sample of observations from the joint posterior distribution $[\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{\Omega}|\mathbf{D}]$, say $\{(\boldsymbol{\theta}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{Z}^{(j)}, \boldsymbol{\Xi}^{(j)}, \boldsymbol{\Omega}^{(j)}), j = 1, \cdots, J\}$.

The Bayesian estimate of $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ can be obtained through the sample mean, and the standard error estimates can be obtained through the sample covariance matrix. Moreover, estimates of the various latent variables can be obtained from $\boldsymbol{\Xi}^{(j)}$ and $\boldsymbol{\Omega}^{(j)}$ for $j = 1, \cdots, J$, respectively. The Gibbs sampler iteratively simulates observations from $p(\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{\Omega}, \mathbf{D})$, $p(\boldsymbol{\alpha}, \mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\Xi}, \boldsymbol{\Omega}, \mathbf{D})$, $p(\boldsymbol{\Xi}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\Omega}, \mathbf{D})$, and $p(\boldsymbol{\Omega}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\Xi}, \mathbf{D})$.

We assume that $p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = p(\boldsymbol{\theta})p(\boldsymbol{\alpha})$. Similarly, we assume a noninformative prior for $\boldsymbol{\alpha}$ and the following conjugate prior distributions for $\boldsymbol{\theta}$:

$$\psi_{yt}^{-1} \stackrel{D}{=} Gamma[a_{0yt}, b_{0yt}], \qquad t = 1, \cdots, T,$$

$$\psi_{xtk}^{-1} \stackrel{D}{=} Gamma[a_{0xtk}, b_{0xtk}], \quad k = 1, \cdots, r, \quad t = 1, \cdots, T,$$

$$\psi_{\omega tk}^{-1} \stackrel{D}{=} Gamma[a_{0\omega tk}, b_{0\omega tk}], \quad k = 1, \cdots, q, \quad t = 1, \cdots, T,$$

$$\psi_{utk}^{-1} \stackrel{D}{=} Gamma[a_{0utk}, b_{0utk}], \quad k = 1, \cdots, p, \quad t = 1, \cdots, T,$$

$$[\boldsymbol{\Lambda}_{utk}|\psi_{utk}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0utk}, \psi_{utk}\boldsymbol{\Sigma}_{0utk}], \quad k = 1, \cdots, p, \quad t = 1, \cdots, T,$$

$$\psi_{\delta k}^{-1} \stackrel{D}{=} Gamma[a_{0\delta k}, b_{0\delta k}], \qquad [\boldsymbol{\Lambda}_{\delta k}|\psi_{\delta k}] \stackrel{D}{=} N[\boldsymbol{\Lambda}_{0\delta k}, \psi_{\delta k}\boldsymbol{\Sigma}_{0\delta k}], \qquad k = 1, \cdots, m,$$

$$\boldsymbol{\beta} \stackrel{D}{=} N[\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0], \qquad \text{and} \qquad \boldsymbol{\Phi}^{-1} \stackrel{D}{=} W_{m(r+q)}[\mathbf{R}_0, \rho_0],$$

where the hyperparameters are defined similarly as before.

With the above prior distributions, the full conditional distributions required for implementing the Gibbs sampler are presented in Appendix 8.1.

The data set in this example is obtained from a longitudinal study about cocaine use conducted by the UCLA Center for Advancing Longitudinal Drug Abuse Research.

Various measures were collected from 321 patients who were admitted in 1989 to the West Los Angeles Veterans Affairs Medical Center and who met the DSM III-R criteria for cocaine dependence (Kasarabada *et al.*, 1999).

These patients were assessed at baseline, one year after treatment, two years after treatment, and 12 years after treatment in 2002 ($t_1 = 0$, $t_2 = 1$, $t_3 = 2$, $t_4 = 12$).

There is a large amount of missing data in this data set. For brevity, the missing data are treated as missing at random (MAR).

sample size is small. if try the missing data as missing at non-random, the model will be unidentifiable since there are much more parameters to be estimated.

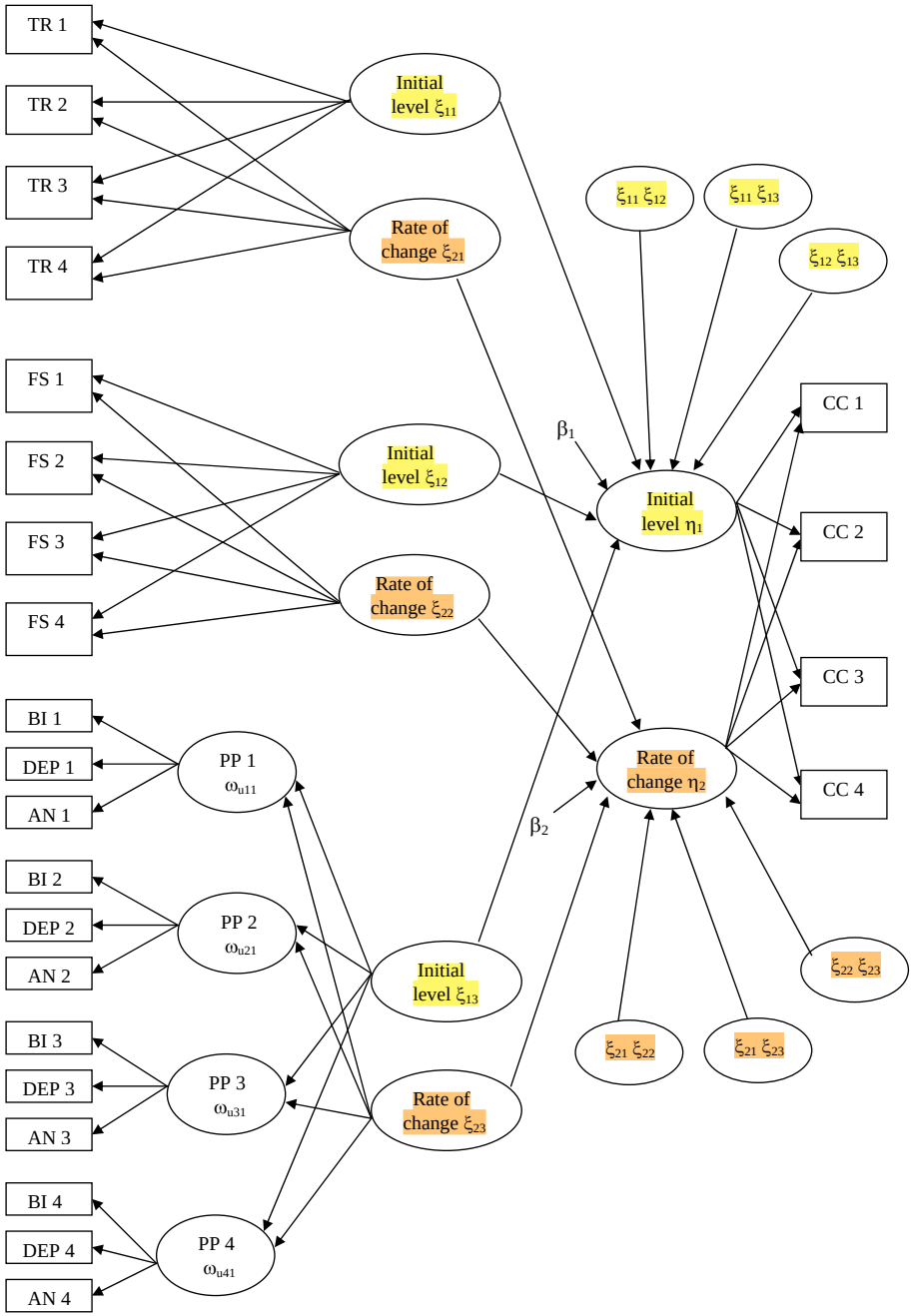
In this analysis, seven variables are involved.

- The first variable 'Days of cocaine use per month (CC)' associates with patients' cocaine use.
- The second variable 'Times per month in formal treatment (TR)' associates with frequencies that the patients participated in treatment.
- The third variable 'Family support (FS)' associates with the family support that the patients had, which is an ordered categorical variable with a 3-point scale.
- Three variables 'Beck Inventory (BI)', 'Depression (DEP)', and 'Anxiety (AN)' associate with the mental health or psychiatric problem of the patients; the latter two are measured based on the Hopkins Symptom Check List (HSCL-58).

In this analysis,

- The outcome latent growth factor (η) represents the pattern of change in CC.
- The first explanatory latent growth factor ξ_1 represents the pattern of change in TR.
- The second explanatory latent growth factor ξ_2 represents the pattern of change in FS.
- The variables BI, DEP, and AN with repeated measurements (at $t = 1, 2, 3, 4$) formed a longitudinal latent variable ω_{ut}, which can be interpreted as 'psychiatric problem (PP)'. Thus, the third explanatory latent growth factors ξ_3 represents the patterns of change in PP.

To study the possible effects of treatment, family support, and psychiatric problem of the patients on cocaine use, we propose the multivariate LCM as described in Figure 8.5.


TR 1
TR 2
TR 3
TR 4
Initial level ξ_{11}
Rate of change ξ_{21}
FS 1
FS 2
FS 3
FS 4
Initial level ξ_{12}
Rate of change ξ_{22}
BI 1
DEP 1
AN 1
PP 1 ω_{u11}
BI 2
DEP 2
AN 2
PP 2 ω_{u21}
BI 3
DEP 3
AN 3
PP 3 ω_{u31}
BI 4
DEP 4
AN 4
PP 4 ω_{u41}
Initial level ξ_{13}
Rate of change ξ_{23}
ξ_{11} ξ_{12}
ξ_{11} ξ_{13}
ξ_{12} ξ_{13}
β_1
Initial level η_1
Rate of change η_2
β_2
CC 1
CC 2
CC 3
CC 4
ξ_{22} ξ_{23}
ξ_{21} ξ_{22}
ξ_{21} ξ_{23}

The interaction effects between the initial status or between the rate of changes are assessed through the structural equations given in (17) and (18).

Among the observed variables, only FS is an ordered categorical variable with a 3-point scale, so we simply denote $\alpha_{tk,j}$ as $\alpha_{t,j}$, and we fix $\alpha_{t,1} = 0.0$ and $\psi_{xt2} = 1.0$ to identify the model.

The Bayesian estimates of the unknown parameters were obtained using WinBUGS and reported in Table 8.1 (see the book chapter). The WinBUGS code is presented in Appendix 8.2.

The estimated structural equations of (17) and (18) are given as follows. The values in parentheses provide the standard error estimates of the associated parameters in the following equations and interpretations.

$$\eta_1 = \underset{(0.254)}{0.566} - \underset{(0.291)}{0.064}\xi_{11} + \underset{(0.362)}{0.383}\xi_{12} + \underset{(0.353)}{0.159}\xi_{13} - \underset{(0.293)}{0.123}\xi_{11}\xi_{12} + \underset{(0.254)}{0.016}\xi_{11}\xi_{13} - \underset{(0.173)}{0.014}\xi_{12}\xi_{13},$$

$$\eta_2 = -\underset{(0.014)}{0.060} + \underset{(0.135)}{0.006}\xi_{21} - \underset{(0.065)}{0.034}\xi_{22} + \underset{(0.219)}{0.517}\xi_{23} + \underset{(0.410)}{0.619}\xi_{21}\xi_{22} + \underset{(0.494)}{0.951}\xi_{21}\xi_{23} + \underset{(0.465)}{0.704}\xi_{22}\xi_{23}.$$

First, $\beta_1 = 0.566$ (0.254), $\beta_2 = -0.060$ (0.014), and $\gamma_{23} = 0.517$ (0.219) are significant. The value of γ_{23} represents the positive effect of rate of change in 'psychiatric problem' on the rate of change in 'cocaine use', suggesting that if a patient's rate of change in 'psychiatric problem' increases over time, his/her rate of change in cocaine use also increases.

Moreover, the following interesting phenomenon is found in predicting the longitudinal change in cocaine use (η_2). Although the coefficients $\gamma_{21} = 0.006$ (0.135), $\gamma_{22} = -0.034$ (0.065) are small and insignificant, the coefficients of interaction terms, $\gamma_{24} = 0.619$ (0.410), $\gamma_{25} = 0.951$ (0.494), and $\gamma_{26} = 0.704$ (0.465), are relatively large. These marginally significant coefficients indicate interactions between the dynamic latent growth factors (ξ_{21} and ξ_{22}, ξ_{21} and ξ_{23}, or ξ_{22} and ξ_{23}), which suggest joint effects of longitudinal changes in treatment, family support, and psychiatric problem on longitudinal change in cocaine use behavior.

This finding shows that the integrated approach (combining formal medical treatment, increasing family support, and reducing psychiatric problem) is more efficient than simple medical treatment in reducing cocaine use of patients.

To better understand the relationships among HRQOL and (MBI, LHS, GDS), we used the LCM to analyze the related longitudinal data.

Ideally, a comprehensive LCM that involves all the twenty-four items associated with HRQOL, MBI, LHS, and GDS should be considered at each of all the time points.

At each time point, physical health (PHY), psychological health (PSY), social interaction (SO), and environment (EN) are identified as latent variables through the measurement equation of an SEM based on the appropriate items in the HRQOL data set. Then the dynamic interrelationships of these latent variables and (MBI, LHS, GDS) are assessed using the comprehensive LCM by considering all the time points together. However, this comprehensive LCM is very tedious.

For brevity, it may be desirable to obtain scores of the abovementioned latent variables first, then the dynamic interrelationships of each of those latent variables with (MBI, LHS, GDS) were analyzed with four separate LCMs, respectively.

There are at least two methods to obtain such scores. One is to use a factor analysis model to estimate the factor scores; while the other is simply to take the averages of the corresponding items. In this illustrative example, we use the later method for simplicity.

Therefore, based on three time points, four LCMs for {PHY, MBI, LHS, GDS}, {PSY, MBI, LHS, GDS}, {SO, MBI, LHS, GDS}, and {EN, MBI, LHS, GDS} are established, respectively. The results of the analysis for the four HRQOL domains were shown in Table 8.3.

As an example, the LCM for physical health domain was shown in Figure 8.6, and the corresponding estimated structural equations, predicted by MBI, LHS, and GDS were as follows:

$$\eta_1 = -0.001 + \underset{(0.069)}{0.174}\xi_1 - \underset{(0.097)}{0.475}\xi_3 - \underset{(0.068)}{0.310}\xi_5,$$

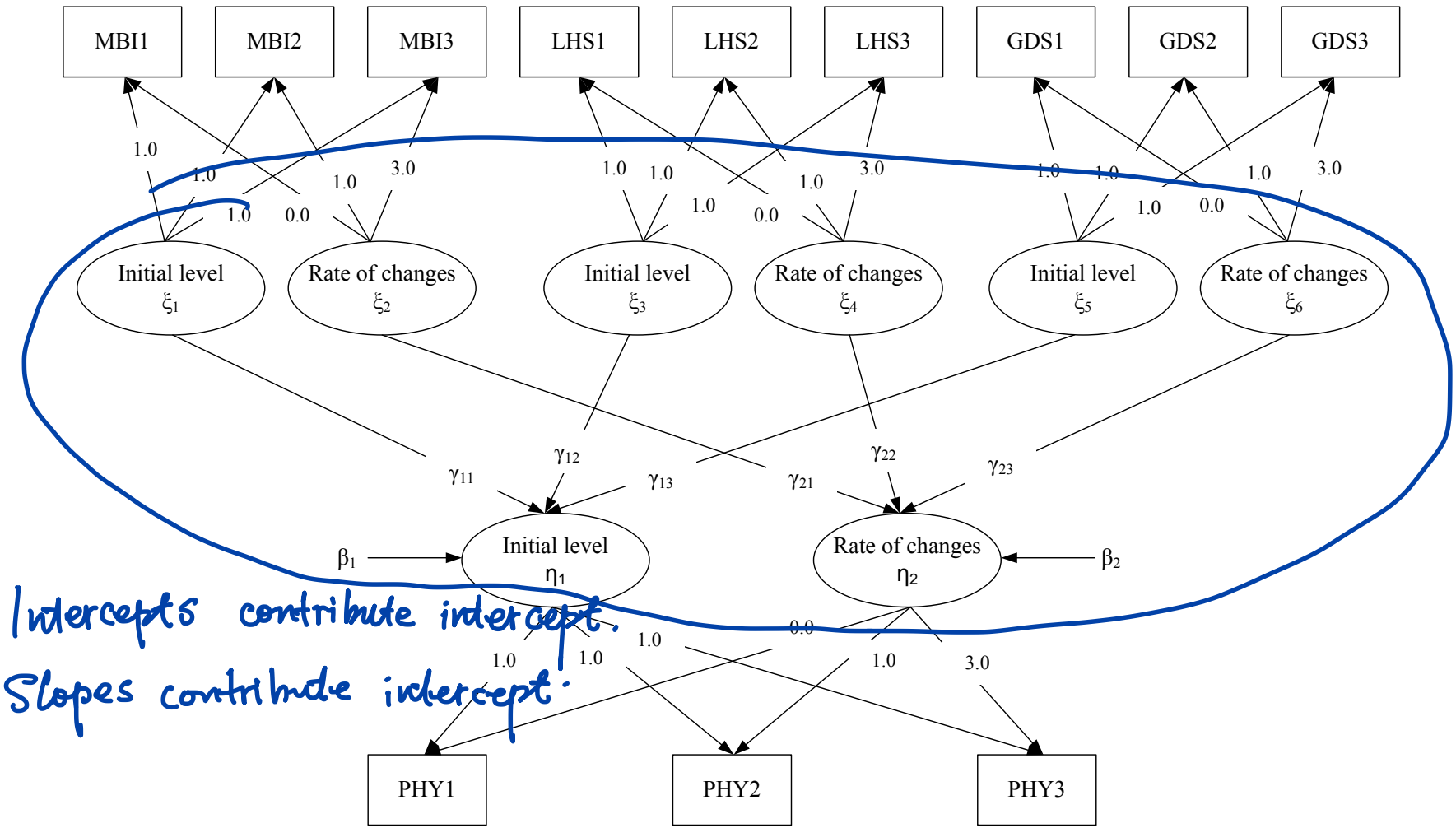
$$\eta_2 = -0.013 + \underset{(0.167)}{0.280}\xi_2 - \underset{(0.163)}{0.312}\xi_4 - \underset{(0.152)}{0.328}\xi_6,$$

where

- η_1 is the initial level of physical health domain,
- η_2 is the rate of change in physical health domain.

Therefore, initial physical health was significantly associated with ADL, handicap, and depression, while the rate of change in physical health was significantly and inversely associated with rate of change in GDS only.

The path diagram of the above model is presented in Figure 8.6.



The initial levels of the other three HRQOL domains were associated with depression only, except that psychological domain was marginally significantly associated with ADL as well. Similar to physical health domain, the rates of change in these three HRQOL domains were associated with the rate of change in depression only.

For the stroke survivors, activities of daily living, handicap, and depression are all important predictors of HRQOL. The relative importance of these interrelated variables in determining HRQOL was the subject of this study. The main finding of our analysis was that change in mood is the most significant effect on the HRQOL of stroke survivors, while changes in basic ADL and handicap had less significant effects; see Pan *et al.* (2008) for more detailed interpretation of the results.

The basic LCM can be generalized to incorporate nonlinear trajectories and effects of covariates. Then, a higher order polynomial is used to describe a nonlinear pattern of dynamic change in individual characteristics.

A longitudinal study about depression in multiple sclerosis (MS) (Beal *et al.*, 2007) is used here to illustrate the nonlinear LCMs. This study focused on three specific research issues:

1. To reveal the patterns of change in depressive symptoms over time.
2. To identify substantial effects of covariates such as age (Age), type of MS (TMS), years since diagnosis of MS (YMS), and functional limitation (FL) on the trajectory of depression over time.
3. To examine the correlations between characteristics of change in functional limitations and depressive symptoms over the seven-year time period.

The data were collected from 607 persons with MS over a seven-year period, with initial recruitment in 1999, as part of an ongoing longitudinal study of quality of life (Stuifbergen *et al.*, 2006) in chronic illness.

A nonlinear trajectory in depression for the sample is suggested by an examination of randomly selected empirical growth plots. A three factor LCM is used to model the trajectories of depression for the MS sample:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_T & t_T^2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{y1} \\ \epsilon_{y2} \\ \vdots \\ \epsilon_{yT} \end{pmatrix}, \tag{20}$$

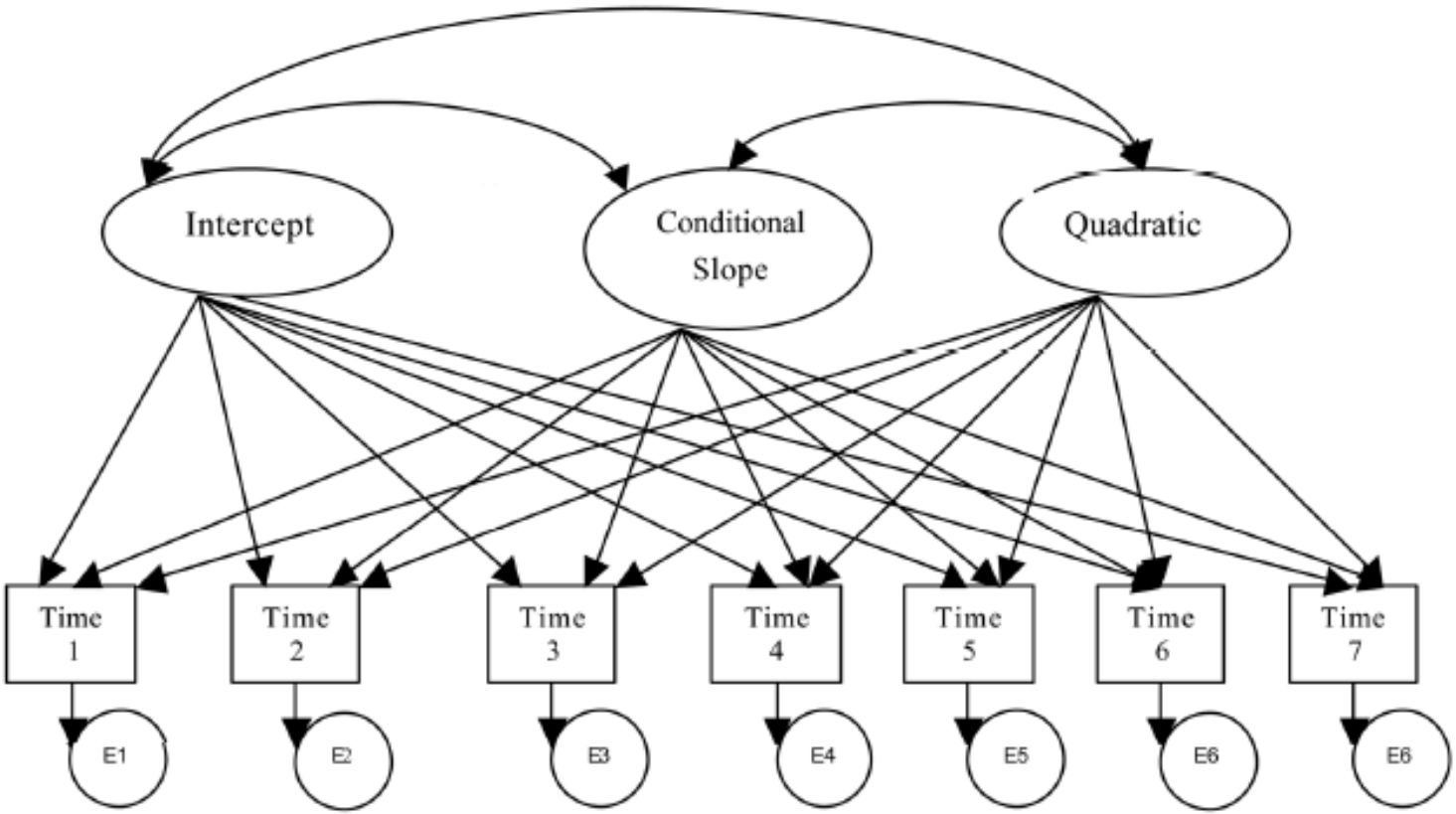
where η_1, η_2, and η_3 represent individual specific random intercept, slope, and quadratic slope, respectively.

$y = f(q) + \varepsilon$

nonparametric model

B/P spline.

The path diagram of model (20) is presented in the following Figure 8.7.



As depicted in Figure 8.7, the change pattern and the correlations between the characteristics of change in depressive symptoms are analyzed.

Furthermore, predictors in change of depression are examined by regressing the intercept, slope, and quadratic slope on the covariates of interest, which leads to the following equations:

$$\eta_1 = \beta_1 + \gamma_{11}(\text{Age}) + \gamma_{12}(\text{TMS}) + \gamma_{13}(\text{FL}) + \delta_1, \tag{21}$$

$$\eta_2 = \beta_2 + \gamma_{21}(\text{Age}) + \gamma_{22}(\text{TMS}) + \gamma_{23}(\text{FL}) + \delta_2, \tag{22}$$

$$\eta_3 = \beta_3 + \gamma_{31}(\text{Age}) + \gamma_{32}(\text{TMS}) + \gamma_{33}(\text{FL}) + \delta_3, \tag{23}$$

Their findings obtained through the nonlinear LCM (Beal *et al.*, 2007) are as follows.

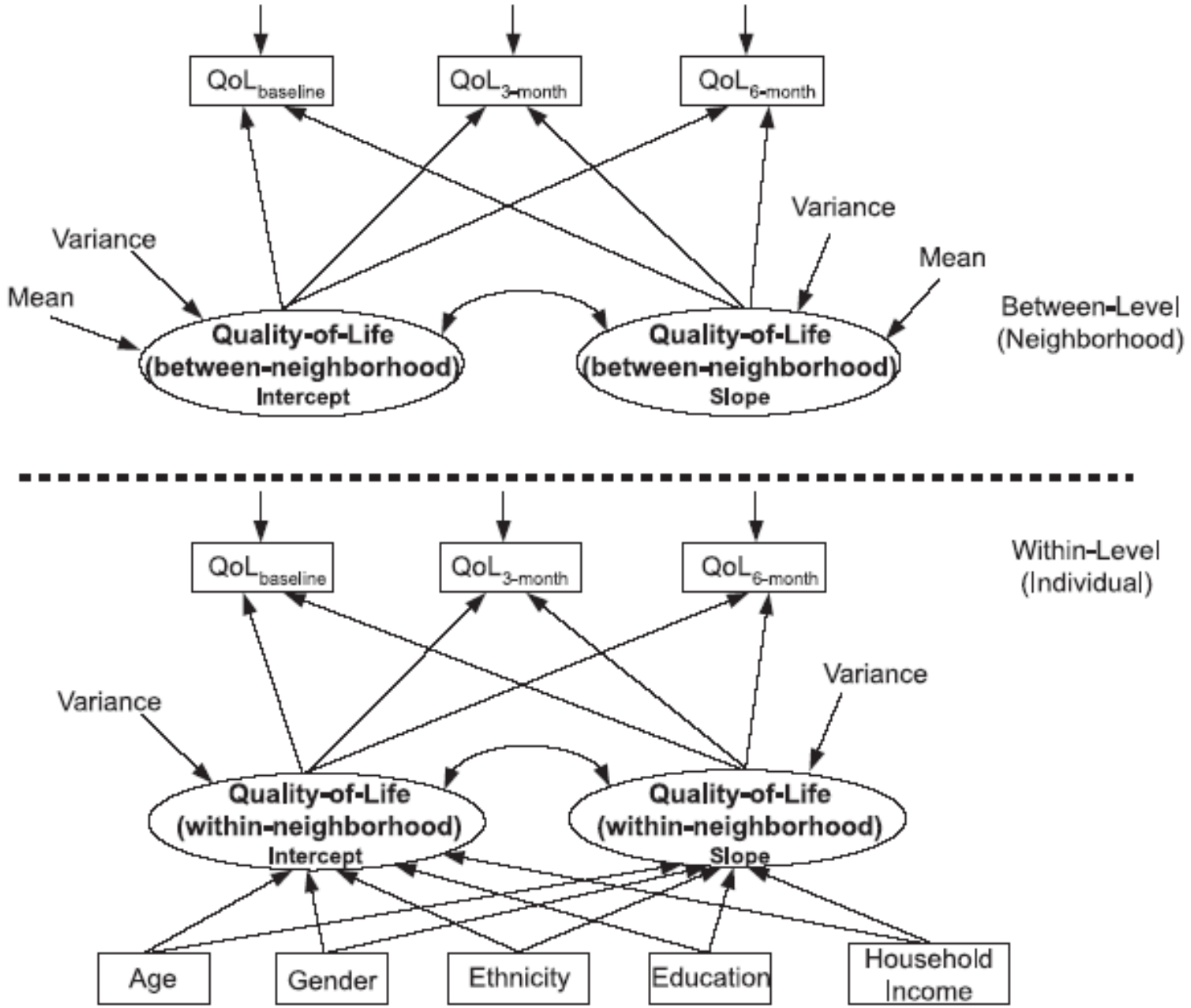
- First, there is no significant increase or decrease trend in depressive symptoms although it fluctuates over time for individuals.
- Second, younger age, longer time since diagnosis of MS, progressive forms of MS, and greater extent of functional limitations will result in greater depressive symptoms at Time 1.
- Third, functional limitation shows an association with depression at all time periods, but other covariates do not. In addition, gender does not predict the changes in depressive symptoms.

These results indicate that screening for depression in all persons with MS is necessary and important.

In some circumstances, clinical trials involve a multilevel design, leading to hierarchically longitudinal data. For example, to evaluate the effects of a neighborhood walking program on quality of life (QOL) among older adults, Fisher and Li (2004) used a multilevel sampling scheme to collect a sample of neighborhoods from a large metropolitan city, from which residents of older adults were randomly recruited. This two-level design results in a nested data structure in which participants are clustered within neighborhoods.

The study focused on whether a 6-month neighborhood walking program improves neighborhood-level QOL for senior residents. A two-level LCM of QOL, with individual- and neighborhood-level data structures, is shown in Figure 8.8. All the involved measures are assessed at baseline, 3 months, and 6 months of the study period.

They concluded that it is feasible and beneficial to implement a neighborhood-based walking program of low to moderate intensity in order to promote QOL among senior residents at a community level.


QoL baseline
QoL 3-month
QoL 6-month
Variance
Mean
Variance
Mean
Quality-of-Life (between-neighborhood) Intercept
Quality-of-Life (between-neighborhood) Slope
Between-Level (Neighborhood)
QoL baseline
QoL 3-month
QoL 6-month
Within-Level (Individual)
Variance
Variance
Quality-of-Life (within-neighborhood) Intercept
Quality-of-Life (within-neighborhood) Slope
Age
Gender
Ethnicity
Education
Household Income

Heterogeneity is commonly encountered in longitudinal analysis of practical applications. For heterogeneous data, there exist some latent classes under which the interested characteristics may present completely different change patterns.

Mixture LCMs can be used to characterize the heterogeneity and reveal specific change pattern for each distinctive latent class. Compared with the basic LCM, the additional tasks in applying mixture LCMs are to identify the number of latent classes to detect the membership of each individual observation and predict the probability of each individual falling in a specific class.

To formulate the probability of individual i belonging to latent class k, the following multinomial logistic regression model is introduced (see also Chapter 7, Section 7.3). For $k = 1, \cdots, K$,

$$p(C_i = k | x_{i1}, \cdots, x_{ip}) = \frac{\exp(a_{0k} + a_{1k}x_{i1} + \cdots + a_{pk}x_{ip})}{\sum_{j=1}^{K} \exp(a_{0j} + a_{1j}x_{i1} + \cdots + a_{pj}x_{ip})}, \quad (24)$$

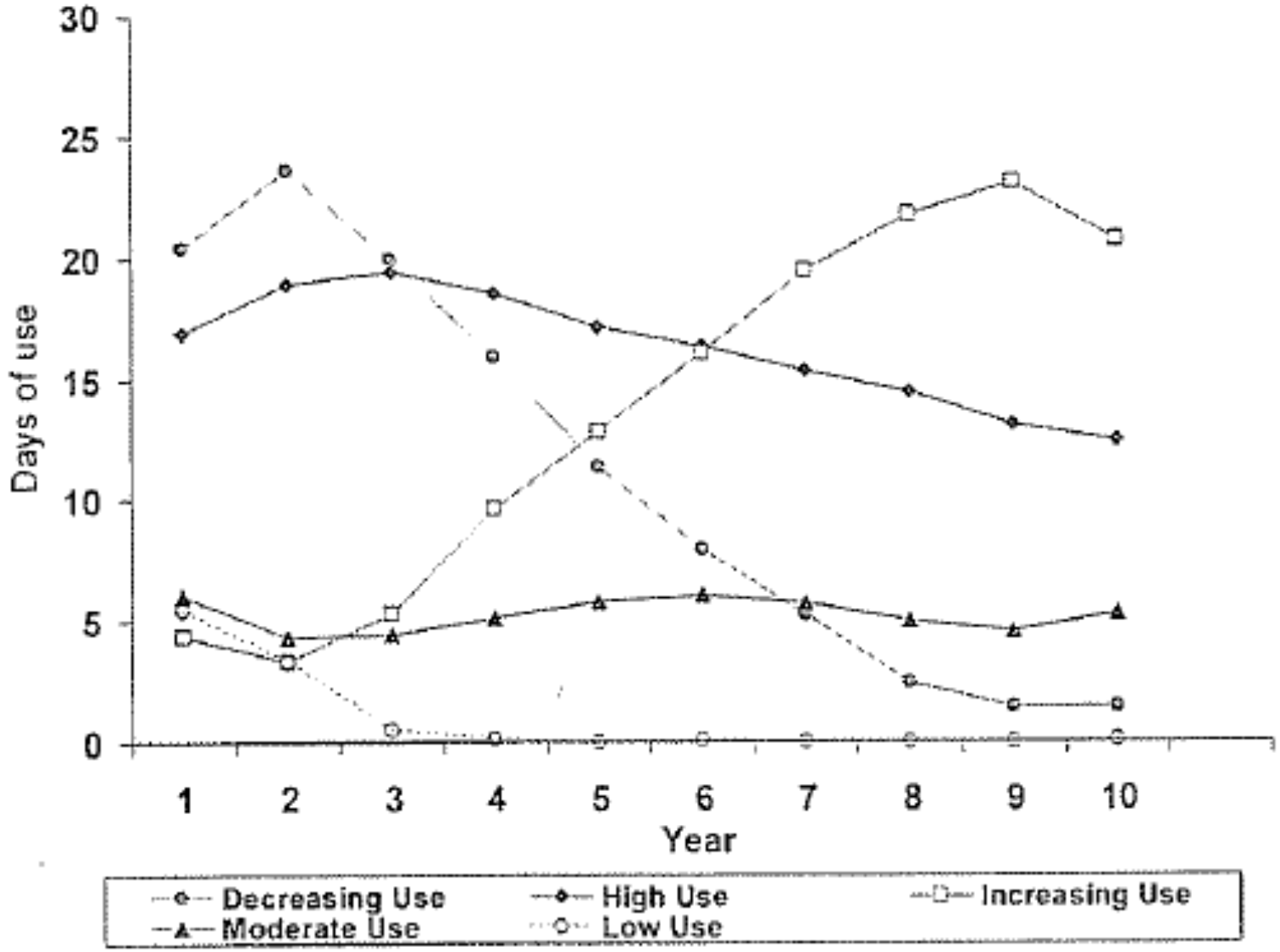
where

- K is the number of latent classes,
- C_i is the class membership for individual i,
- $x_{i1}, \cdots, x_{ip}$ are covariates that may potentially influence the chance of individual i belonging to latent class k,
- $a_{0k}, a_{1k}, \cdots, a_{pk}$ are corresponding regression coefficients that reflect the importance of potential covariates.

Hser *et al.* (2008) applied this model to examine long-term trajectories of drug use for primary heroin, cocaine, and meth users. The data include 629 primary heroin users, 694 cocaine users, and 474 meth users. The main outcome measure is the number of days using the primary drug per month.

Figure 8.9 shows that the analysis of mixture LCMs reveals five distinctive groups with different drug use trajectories over ten-year follow-up: consistently high use, increasing use, decreasing use, moderate use, and low use. The primary drug type is significantly associated with different trajectory patterns. Heroin users are most likely in the consistently high use group, cocaine and meth users are most likely in the moderate use group.

Moreover, the high use group have earlier onsets of drug use and crime, longer incarceration durations, and less employed periods than those in other groups. Compared with other existing studies of drug addiction, the use of mixture LCMs emphasizes the heterogeneity of drug use patterns and the importance of understanding and addressing the full spectrum of drug use patterns over time.


30
25
20
15
10
5
0
Days of use
1
2
3
4
5
6
7
8
9
10
Year
Decreasing Use
High Use
Increasing Use
Moderate Use
Low Use

Another application of mixture LCMs is the analysis of depression in the persons with myocardial infarction (MI). Elliott *et al.* (2006) analyzed affect and event data from subjects post-MI in order to understand how mood and reactivity to negative events over time relate to diagnostic level depression.

In this study, 35 patients who had experienced an MI within the past year and were in treatment currently are investigated. The affect scores and event indicators (indicating presence of positive, negative, and neutral events) of the patients are collected for up to 35 consecutive days.

The analysis of mixture LCMs suggests a two-class model for the MI patients: an optimist class with stable positive affect and declining perceived negative events; and a pessimist class with declining positive affect and continuing perceived negative events.

Depressed subjects have a 92% chance of belonging to the pessimist class, compared with 62% among non-depressed subjects. This finding uncovers some hitherto unobserved structure in the positive affect and negative event data in this sample.

The key advantage of using mixture LCMs in this study is that persons who are most at risk of developing major or minor depression can be identified, which will assist in developing more specific interventions or treatments.


Ahlsiö, B., Britton, M., Murray, V. and Theorell, T. (1984) Disablement and quality of life after stroke. *Stroke*, **15**, 886-890.


Beal, C. C., Stuifbergen, A. K., Sands, D. V. and Brown, A. (2007) Depression in multiple sclerosis: a longitudinal analysis. *Archives of Psychiatric Nursing*, **21**, 181-191.


Bollen, K. A. and Curran, P. J. (2006) *Latent Curve Models - A Structural Equation Perspective*, New Jersey: John Wiley & Sons, Inc.


Brown, R. A., Monti, P. M., Myers, M. G., Martin, R. A., Rivinus, T., Dubreuil, M. E. T. and Rohsenow, D. J. (1998) Depression among cocaine abusers in treatment: relation to cocaine and alcohol use and treatment outcome. *The American Journal of Psychiatry*, **155**, 220-225.


Elliott, M. R., Gallo, J. J., Ten Have, T. R., Bogner, H. R. and Katz, I. R. (2006) Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction. *Biostatistics*, **6**, 119-143.


Fisher, K. J. and Li, F. Z. (2004) A community-based walking trial to improve neighborhood quality of life in older adults: a multilevel analysis. *Annals of Behavioral Medicine*, **28**, 186-194.


Hser, Y. I., Huang, D., Brecht, M. L., Li, L. B. and Evans, E. (2008) Contrasting trajectories of heroin, cocaine, and methamphetamine. *Journal of Addictive Diseases*, **27**, 13-21.


Hser, Y. I., Stark, M. E., Paredes, A., Huang, D., Anglin, M. D. and Rawson, R. (2006) A 12-year follow-up of a treated cocaine-dependent sample. *Journal of Substance Abuse Treatment*, **30**, 219-226.


Jöreskog, K. G. (1970) A general method for analysis of covariance structures. *Biometrika*, **57**, 239-251.


Kasarabada, D. N., Anglin, M. D., Khalsa-Denison, E. and Paredes, A. (1999) Differential effects of treatment modality on psychosocial functioning of cocaine-dependent men. *Journal of Clinical Psychology*, **55**, 257-274.


Kwok, T., Lo, R. S., Wong, E., Wai-Kwong, T., Mok, V. and Kai-Sing, W. (2006) Quality of life of stroke survivors: a 1-year follow-up study. *Archives of Physical Medicine and Rehabilitation*, **87**, 1177-1182.


Meredith, W. and Tisak, J. (1990) Latent curve analysis. *Psychometrika*, **55**, 107-122.


Pan, J. H., Song, X. Y., Lee, S. Y. and Kwok, T. (2008) Longitudinal analysis of quality of life for stroke survivors using latent curve models. *Stroke*, **39**, 2795-2802.


Song, X. Y., Lee, S. Y. and Hser, Y. I. (2009) Bayesian analysis of multivariate latent curve models with nonlinear longitudinal latent effects. *Structural Equation Modeling - A Multidisciplinary Journal*, **16**, 245-266.


Stuifbergen, A. K., Blozis, S. A., Harrison, T. C. and Becker, H. A. (2006) Exercise, functional limitations, and quality of life: A longitudinal study of persons with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation*, **87**, 935-943.