A Latent Gaussian Process Model for Analyzing Intensive Longitudinal Data

Yunxiao Chen and Siliang Zhang

Abstract

Intensive longitudinal studies are becoming progressively more prevalent across many social science areas, especially in psychology. New technologies like smartphones, fitness trackers, and the Internet of Things make it much easier than in the past for data collection in intensive longitudinal studies, providing an opportunity to look deep into the underlying characteristics of individuals under a high temporal resolution. In this paper, we introduce a new modeling framework for latent curve analysis that is more suitable for the analysis of intensive longitudinal data than existing latent curve models. Specifically, through the modeling of an individual-specific continuous-time latent process, some unique features of intensive longitudinal data are better captured, including intensive measurements in time and unequally spaced time points of observations. Technically, the continuous-time latent process is modeled by a Gaussian process model. This model can be regarded as a semi-parametric extension of the classical latent curve models and falls under the framework of structural equation modeling. Procedures for parameter estimation and statistical inference are provided under an empirical Bayes framework and evaluated by simulation studies. We illustrate the use of the proposed model though the analysis of an ecological momentary assessment dataset.

KEY WORDS: Gaussian process, latent curve analysis, structural equation modeling, intensive longitudinal data, ecological momentary assessment, time-varying latent trait

1 Introduction

Intensive longitudinal data are becoming progressively more prevalent across many social science areas, especially in psychology, catalysed by technological advances (e.g., Chapter 1, Bolger & Laurenceau, 2013). Such data usually involve many repeated measurements that reflect individual-specific change process in high resolution, enabling researchers to answer deeper research questions of human behavioral patterns. Due to the complex structure of intensive longitudinal data, statistical models play an important role in the analysis of such data.

In an intensive longitudinal study, repeated measurements are made intensively over time. Such data may involve (1) a large number of time points, (2) individually-varying numbers of observations, (3) unequally spaced time points of observations, and (4) response data of various types (e.g., continuous, ordinal, etc.). For example, consider intensive longitudinal data from ecological momentary assessment (EMA) under a signal-contingent sampling scheme (see Chapter 5, Conner & Lehman, 2012), which repeatedly measures individuals' current behaviors and experiences in real time, in the individuals' natural environments. Under this sampling scheme, participants are "beeped" at several (random) times a day to complete an electronic diary record on psychological variables, such as symptoms or well-being. The assessments can last for many days (e.g. a month). Such a design has been used to study, for example, borderline personality disorder (Trull et al., 2008), adolescent smoking (Hedeker et al., 2012), and others. We visualize this design in Figure 1, where the measurements happen at time points marked by "$\mathbf{x}$". Under such a design, each individual may receive hundreds of repeated measurements at

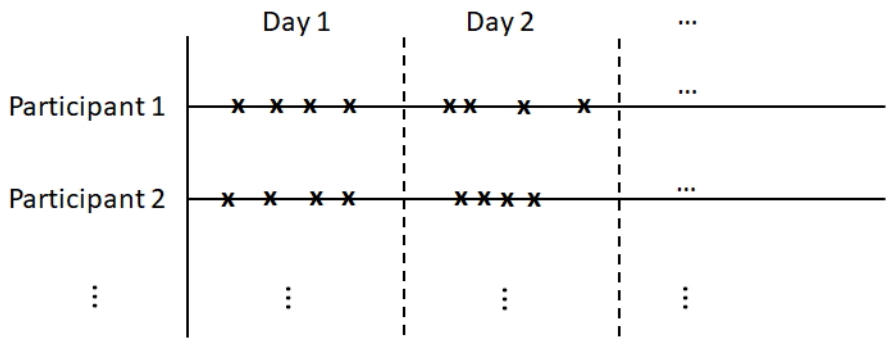


Figure 1: An illustration of the signal-contingent sampling scheme of an ecological momentary assessment.

irregularly spaced time points. Depending on the measurement scale, one or multiple indicators may be recorded at each observation time point and the indicators can be either continuous or categorical.

Latent curve models (e.g. Bollen & Curran, 2006; Duncan et al., 2013; Ram & Grimm, 2015), also known as latent growth models or growth curve models, are an important family of psychometric models for the analysis of longitudinal measurements. These models characterize the growth or change in an individual through the modeling of an individual-specific time-varying latent trait, where the latent trait often has a substantive interpretation, such as a cognitive ability, a psychopathological trait, or subjective well-being. Such models are typically formulated under the structural equation modeling framework. In these models, each individual i is represented by a latent curve $\{\theta_i(t) : t \geq 0\}$, which represents a time-varying latent trait. At a given observation time t, the individual's response to a single or multiple items is assumed to be driven by his/her current latent trait level $\theta_i(t)$.

The classical latent curve models are developed for non-intensive longitudinal data (typically less than 10 times of measurement). Therefore, they often make strong assumptions on the functional form of $\theta_i(t)$. For example, a linear latent curve model assumes that $\theta_i(t) = \beta_{i0} + \beta_{i1}t$, where β_{i0} and β_{i1} are the intercept and the slope of the

curve, treated as individual specific latent variables. In other words, in this linear curve model, the latent curve $\theta_i(t)$ is a random function, characterized by two random effects β_{i0} and β_{i1} that are often assumed to follow a bivariate normal distribution. Although $\theta_i(t)$ can take slightly more complex forms (e.g., polynomial), the functional form of $\theta_i(t)$ in the classical models is usually simple, which may not be suitable for analyzing individual change processes revealed by intensive longitudinal data, where the number of measurements may vary across different individuals.

To better capture the temporal pattern in intensive longitudinal data, more flexible latent curve models have been proposed under the structural equation modeling framework. Depending on whether time is treated as discrete or continuous, these models can be classified into two categories. The discrete-time models are typically a hybrid of time series analysis models and the structural equation modeling framework. Specifically, the individual specific dynamic latent traits are modeled by a time series model, such as the autoregressive (AR) or vector autoregressive (VAR) models. Such models are usually known as the latent variable-autoregressive latent trajectory models (Bianconcini & Bollen, 2018) or dynamic structural equation models (Asparouhov et al., 2018). The continuous-time models typically assume that the dynamic latent traits follow a stochastic differential equation (SDE; Oud & Jansen, 2000; Voelkle et al., 2012; Lu et al., 2015). For example, Lu et al. (2015) assume the dynamic latent trait to follow the Ornstein-Uhlenbeck Gaussian process (Uhlenbeck & Ornstein, 1930), whose distribution is given by an SDE.

The above models have limitations. Discrete-time models may be over-simplified for intensive longitudinal data, for which measurement occurs in continuous time. In particular, when time points of measurements are irregularly spaced and different individuals have different numbers of measurements, it is difficult to organize intensive longitudinal data into the format of multivariate time-series data and then analyze using a discrete-

time model. Arbitrarily transforming data into a multivariate time-series format is likely to introduce bias into the analysis, as time lags between measurements, which may vary substantially among individuals, are ignored in the discrete-time formatting. In theory, these issues with discrete-time models can be addressed by taking a continuous-time model. However, existing continuous-time models are typically not straightforward to specify, estimate, and make inference upon, as latent stochastic differential equations are not straightforward to deal with either analytically or numerically. Moreover, limited by the form of stochastic differential equations, the existing continuous-time models for insensitive longitudinal data may not be rich enough.

In this paper, we propose new continuous-time latent curve models for the analysis of intensive longitudinal data that do not suffer from the issues with the existing models and better capture the unique features of intensive longitudinal data mentioned previously. By imposing Gaussian process models (Rasmussen & Williams, 2005) on the latent curves $\{\theta_i(t) : t \geq 0\}$, a general framework for latent curve modeling is developed. We call it the *Latent Gaussian Process* (LGP) models. In contrast to discrete-time models, the proposed models retain the flexibility of continuous-time models in dealing with observations in a continuous time domain. In addition, this general framework contains models that are easier to specify and analyze than SDE-based models.

Technically, the proposed modeling framework can be viewed as a hybrid of the latent Gaussian process model for functional data analysis (Hall et al., 2008) and the generalized multilevel structural equation modeling framework for longitudinal measurement (e.g., Chapter 4, Skrondal & Rabe-Hesketh, 2004). As will be shown in the sequel, many existing latent curve models, whether time is treated as continuous or discrete, can be viewed as special cases under the proposed general framework. By making use of mathematical characterizations of Gaussian processes, methods for the parametrization of LGP models are provided. In addition, parameter estimation and statistical inference

are carried out under an empirical Bayes framework, using a Stochastic Expectation-Maximization (StEM) algorithm (Celeux & Diebolt, 1985; Nielsen, 2000; Zhang et al., 2018).

The rest of the paper is organized as follows. In Section 2, the classical latent curve models are reviewed under a unified framework of structural equation modeling and then a new latent Gaussian process modeling framework is introduced that substantially generalizes the traditional models. The parametrization of latent Gaussian process models is discussed. Estimation and statistical inference are discussed in Section 3, followed by the computational details in Section 4. Extension to the incorporation of covariates is discussed in Section 5. The proposed model is evaluated in Section 6 through simulation studies and further illustrated in Section 7 via a real data example. We end with concluding remarks in Section 8.

2 Latent Gaussian Process Model

2.1 A Unified Framework for Latent Curve Analysis

We first provide a unified framework for latent curve analysis. We consider N participants being measured longitudinally within a time interval $[0, T]$, where time is treated as continuous. For individual i, let $t_{is} \in [0, T]$ be the time that the sth measurement occurs and S_i be the total number of measurements received by individual i. At each time $t = t_{i1}, ..., t_{iS_i}$, we observe a random vector $\mathbf{Y}_i(t) = (Y_{i1}(t), ..., Y_{iJ}(t))^\top$, where $Y_{ij}(t)$ can be either continuous or categorical, depending on the data type of the jth indicator. In particular, the corresponding latent curve model is called a single-indicator model when $J = 1$ and a multiple-indicator model when $J > 1$. We denote $\mathbf{y}_i(t) = (y_{i1}(t), ..., y_{iJ}(t))^\top$ as a realization of $\mathbf{Y}_i(t)$. Moreover, each individual i is associated with a latent curve $\theta_i(\cdot) = \{\theta_i(t) : t \in [0, T]\}$, which can be regarded as a time-varying latent trait. Note

that the above setting is quite general that includes discrete-time longitudinal data as a special case, for which the observation time t_{is} takes value in $\{0, 1, 2, ...\}$.

The latent curve model consists of two components: (1) a measurement model that specifies the conditional distribution of $\{\mathbf{Y}_i(t) : t = t_{i1}, ..., t_{iS_i}\}$ given $\{\theta_i(t) : t \in [0, T]\}$, and (2) a structural model that specifies the distribution of the random function $(\theta_i(t) : t \in [0, T])$.

Measurement model. The measurement model assumes that the distribution of $\mathbf{Y}_i(t)$ only depends on $\theta_i(t)$, the latent trait level at the same time point, but does not depend on the latent trait levels or responses at any other time points. More precisely, it is assumed that

$$f(\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})|\theta_i(t), t \in [0, T]) = f(\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})|\theta_i(t_{i1}), ..., \theta_i(t_{iS_i})), \qquad (1)$$

where $f(\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})|\theta_i(t), t \in [0, T])$ denotes the probability density/mass function of the conditional distribution of $\mathbf{Y}_i(t_{i1})$, ..., $\mathbf{Y}_i(t_{iS_i})$ given the entire latent process $(\theta_i(t) : t \in [0, T])$ and $f(\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})|\theta_i(t_{i1}), ..., \theta_i(t_{iS_i}))$ denotes the probability density/mass function of the conditional distribution of $\mathbf{Y}_i(t_{i1})$, ..., $\mathbf{Y}_i(t_{iS_i})$ given $\theta_i(t_{i1}), ..., \theta_i(t_{iS_i})$. Equation (1) means that the latent trait level at any other time point is conditionally independent of the observed responses, given the latent trait levels at the corresponding time points of observation. As visualized in Figure 2, it is further assumed that the conditional distribution (1) has the following decomposition,

$$f(\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})|\theta_i(t_{i1}), ..., \theta_i(t_{iS_i})) = \prod_{s=1}^{S_i} g(\mathbf{y}_i(t_{is})|\theta_i(t_{is})), \qquad (2)$$

where $g(\mathbf{y}_i(t)|\theta_i(t))$ is the conditional probability density/mass function of $\mathbf{Y}_i(t)$ given $\theta_i(t)$. The assumption in (2) is conceptually similar to the widely used local independence

assumption in latent variable models (see Chapter 4, Skrondal & Rabe-Hesketh, 2004). Finally, we assume local independence among multiple indicators at each time t, i.e., $Y_{i1}(t)$, ..., $Y_{iJ}(t)$ are conditionally independent given $\theta_i(t)$. That is

$$g(\mathbf{y}_i(t)|\theta_i(t)) = \prod_{j=1}^{J} g_j(y_{ij}(t)|\theta_i(t)), \tag{3}$$

where $g_j(y_{ij}(t)|\theta_i(t))$ specifies the conditional distribution of the jth indicator $Y_{ij}(t)$ given $\theta_i(t)$. The choice of g_j depends on the type of the jth indicator. It is worth noting that the conditional distribution g_j does not depend on time t, implying that the measurement is assumed to be time-invariant. Although commonly adopted in latent curve models (e.g., Chapter 2, Bollen & Curran, 2006), this assumption is quite strong and needs to be checked when applying such models to real data.

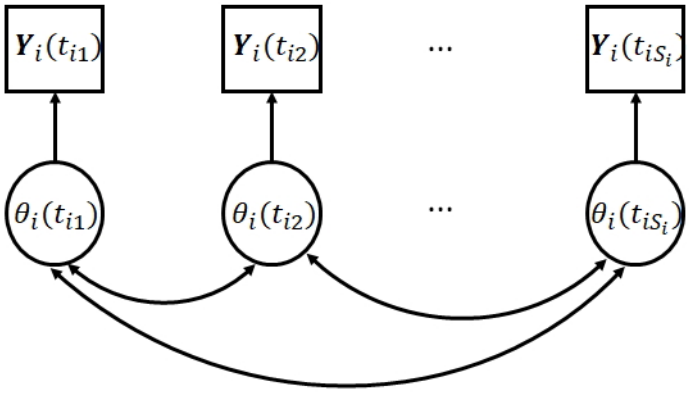


Figure 2: Path diagram for a unified latent curve model.

We provide several measurement model examples.

1. Linear factor model for continuous response:

$$Y_{ij}(t)|\theta_i(t) \sim N(a_j\theta_i(t) + b_j, \sigma_j^2), \tag{4}$$

where a_j, b_j, and σ_j^2 are model parameters.

2. Probit model for ordinal response ($Y_{ij} \in \{0, 1, ..., n_j\}$):

$$P(Y_{ij}(t) = l|\theta_i(t)) = \Phi\left(b_{j,l+1} + a_j\theta_i(t)\right) - \Phi\left(b_{j,l} + a_j\theta_i(t)\right), \tag{5}$$

where

$$-\infty = b_{j,0} < b_{j,1} < b_{j,2} < \ldots < b_{j,n_j} < b_{j,n_j+1} = \infty.$$

$b_{j,l}$ and a_j are model parameters, $l \in \{1, ..., n_j\}$ and $j = 1, ..., J$. When $n_j = 1$, Y_{ij} degenerates to a binary response variable and the model (5) becomes the well-known two-parameter normal-ogive model in item response theory (Chapter 4, Embretson & Reise, 2000).

Model (5) can be specified alternatively through the introduction of latent responses. That is, define latent response

$$Y^*_{ij}(t) = -a_j\theta_i(t) + \epsilon_{ij}(t)$$

where $\epsilon_{ij}(t)$ is a noise term following a standard normal distribution. Then the observable response $Y_{ij}(t)$ can be viewed as a truncated version of $Y^*_{ij}(t)$, obtained by

$$Y_{ij}(t) = l \text{ if } b_{j,l} \leq Y^*_{ij}(t) < b_{j,l+1}.$$

When the multiple indicators contain a mixture of ordinal and continuous variables, the above models can be combined to model $Y_{i1}(t), ..., Y_{iJ}(t)$, since the measurement models for different items can be specified independently given the local independence assumption.

Structural model. The structural model specifies the distribution of the random function $\theta_i(t)$. We list a few examples below and refer the readers to Bollen & Curran

(2006) for a comprehensive review.

1. Linear trajectory model:

$$\theta_i(t) = \beta_{i0} + \beta_{i1}t, \tag{6}$$

where $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1})$ are individual specific random effects, following a bivariate normal distribution.

2. Quadratic trajectory model:

$$\theta_i(t) = \beta_{i0} + \beta_{i1}t + \beta_{i2}t^2, \tag{7}$$

where $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \beta_{i2})$ are individual specific random effects, following a trivariate normal distribution.

3. Exponential trajectory model:

$$\theta_i(t) = \beta_{i0} + \beta_{i1}\exp(\gamma t), \tag{8}$$

where $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1})$ are individual specific random effects, following a bivariate normal distribution and γ is a fixed effect parameter.

These models assume a simple functional form for $\theta_i(t)$. In particular, the realizations of $\theta_i(t)$ are restricted to linear, quadratic, and exponential functions for models (6)-(8), respectively. Such models tend to be effective for non-intensive longitudinal data (typically less than 10 measurements), but may not be flexible enough when having intensive longitudinal measurements which provide information in a high temporal resolution. In the rest of the paper, a general modeling framework is proposed, based on which more flexible structural models can be constructed.

2.2 Gaussian Process Structural Model

In what follows, we introduce a new framework for modeling $\theta_i(t)$ as a continuous-time stochastic process. A key component of this framework is the Gaussian process model.

Definition 1 (Gaussian Process) *A time continuous stochastic process $X(t)$ on time interval $[0,T]$ is a Gaussian process if and only if for every finite set of time points $t_1, ..., t_S \in [0,T]$, $(X(t_1), ..., X(t_S))$ is multivariate normal.*

We remark that a Gaussian process can be defined more generally on a real line. In this paper, we focus on Gaussian process on a bounded interval $[0,T]$, since real longitudinal data are collected within a certain time window. Many widely used stochastic processes, including the Brownian motion, the Brownian bridge, and the Ornstein-Uhlenbeck process, are special cases of Gaussian process. Thanks to the flexibility, nonlinearity, and inherent nonparametric structure, Gaussian processes have been widely used as a model for random functions for solving regression, classification, and dimension reduction problems (Chapter 4, Rasmussen & Williams, 2005).

Thanks to the normality, a Gaussian process is completely characterized by two components: (1) a mean function $m(t) = EX(t)$, and (2) a kernel function $K(t,t')$ for the covariance structure, where $K(t,t') = Cov(X(t), X(t'))$. We provide a definition of a kernel function below.

Definition 2 (Kernel Function) *A bivariate function $K(t,t')$ is called a kernel function if for every finite set of points $t_1, ..., t_S$, the matrix $(K(t_i,t_j) : i,j = 1, ..., S)$ is positive semidefinite.*

Note that since $K(t,t') = Cov(X(t), X(t'))$, the matrix $(K(t_i,t_j) : i,j = 1, ..., S)$ has to be positive semidefinite, because it is the covariance matrix of $(X(t_1), ..., X(t_S))$. On the other hand, it can be shown that for any kernel function K, there exists a Gaussian process whose covariance structure is given by the kernel (Chapter 4, Rasmussen &

Williams, 2005). As an illustrative example, Figure 3 shows three independent realizations from a Gaussian process, with a mean function $m(t) = 0$ and a squared exponential kernel function $K(t, t') = \exp(-(t - t')^2/(2 \times 0.5^2))$.

Definition 3 (Gaussian Process Structural Model) *We say the structural component of a latent curve model follows a Gaussian process structural model, if* $\{\theta_i(t) : t \in [0, T]\}$ *are independent and identically distributed (i.i.d.) Gaussian processes for* $i = 1, ..., N$.

We remark that the Gaussian process structural model assumption in Definition 3 can be viewed as an extension of a commonly adopted assumption in unidimensional or multidimensional item response theory models where individual-specific latent trait or traits are assumed to be i.i.d. univariate or multivariate normal. The difference is that, rather than having a random variable or random vector for each individual, each individual in the proposed model is characterized by a random function, whose distribution is less straightforward to parameterize.

Combining a Gaussian process structural model and a measurement model as defined in Section 2.1, we obtain an LGP model. We point out that the examples (6)-(8) are all special cases of the LGP model. This is because, due to the multivariate normality of the random effects, for every finite set of time points $t_1, ..., t_S$, $(\theta_i(t_1), ..., \theta_i(t_S))$ is multivariate normal. In addition, all the SDE based continuous-time latent curve models also fall into this framework, when the noise component of the SDE is assumed to be Gaussian. For example, Lu et al. (2015) assume the dynamic latent trait to follow the Ornstein-Uhlenbeck process (Uhlenbeck & Ornstein, 1930). This process is a Gaussian process described by a stochastic differential equation with Gaussian noise. Furthermore, when the latent variables are assumed to be jointly normal, the latent variable-autoregressive latent trajectory models (see Bianconcini & Bollen, 2018), which are discrete-time mod-

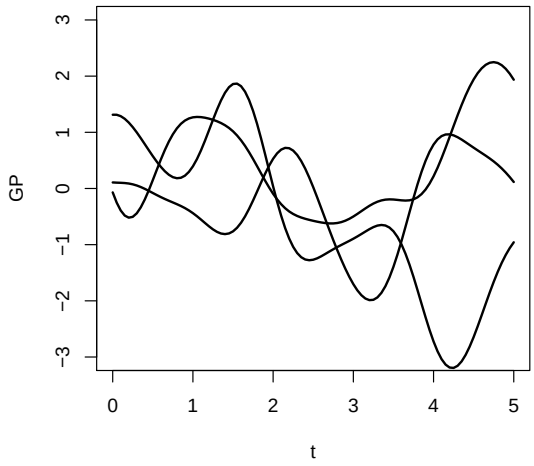


Figure 3: Sample paths from a Gaussian process, where $m(t) = 0$ and $K(t, t') = \exp\left(-\frac{(t-t')^2}{2\times 0.5^2}\right)$

els, can also be viewed as special cases under the current framework.

A Gaussian process is specified by a mean function $m(t)$ and a kernel function $K(t, t')$, whose choices should be problem specific. We denote the distribution of such a stochastic process by $\mathrm{GP}(m, K)$. In what follows, we discuss the parametrization of Gaussian process structural models.

2.3 Parametrization of Gaussian Process Structural Model

Following the above discussion, we see that $\theta_i(t) = m(t) + \bar{\theta}_i(t)$, where $\bar{\theta}_i(t)$ is Gaussian process with mean 0 and kernel $K(t, t')$. This allows us to discuss the modeling of $m(t)$ and $K(t, t')$ separately, while in the classical latent curve models (e.g., (6)-(8)) the mean and kernel are modeled simultaneously. In particular, the mean process $m(t)$ can be viewed as the mean of $\theta_i(t)$, for individuals from a population of interest. Therefore, the mean function captures the mean level of the time-varying latent trait, possibly reflecting the trend and the periodicity of the dynamic latent trait at the population

level. In addition, the mean zero Gaussian process $\bar{\theta}_i(t)$ can be viewed as the deviation from the mean process that is specific to individual i.

Mean function. We consider the parametrization of the mean function $m(t)$, which is typically assumed to have certain level of smoothness. Specifically, for the linear, the quadratic, and the exponential trajectory models mentioned in Section 2.1, the mean functions $m(t)$ take linear, quadratic, and exponential forms.

Under the current framework, $m(t)$ can be parameterized more flexibly. Specifically, we adopt a parametrization of $m(t)$ using basis functions. That is,

$$m(t) = \alpha_0 + \alpha_1 b_1(t) + \cdots + \alpha_D b_D(t), \tag{9}$$

where $b_1(t)$, ..., $b_D(t)$ are pre-specified basis functions on $[0, T]$. For example, when polynomial basis functions are used, $b_d(t) = t^d$, $d = 1, 2, ..., D$, where D is the degree of the polynomial function. When cubic spline basis functions are used, $b_1(t) = t$, $b_2(t) = t^2$, $b_3(t) = t^3$, and $b_{3+d} = (t - \xi_d)_+^3$, where $d = 1, ..., D - 3$, ξ_d is the dth spline knot that is pre-specified on $[0, T]$, and $(t-\xi_d)_+^3 = (t-\xi_d)^3$ when $t > \xi_d$ and 0 otherwise. Alternative basis functions may also be used, such as Fourier basis, wavelets, and other spline basis functions. We refer the readers to Chapter 3, Ramsay & Silverman (1997) for a review of different basis functions. We remark that the number of basis functions D and the choices of basis function may be determined by data through model comparison.

We remark that if the dynamic trait $\theta_i(t)$ is assume to be a stationary process (i.e., the joint distribution of $\theta_i(t)$ does not change when the process is shifted in time), then the mean function does not depend on time t. In that case, the mean function can only have an intercept parameter, $m(t) = \alpha_0$.

Parameterizing kernel function. One way to model the mean zero Gaussian process $\bar{\theta}_i(t)$ is by directly parameterizing the kernel function. In fact, different parametric kernel functions are available in the literature. We refer the readers to (Chapter 4, Rasmussen & Williams, 2005) for a review. In what follows, we provide a few examples of kernel functions, with a focus on kernels that lead to stationary mean zero Gaussian processes. For such a kernel function $K(t, t')$, the value of $K(t, t')$ only depends on the time lag $|t - t'|$, not the specific values of t and t'. A stationary kernel should be used if the distribution of $\bar{\theta}_i(t)$ is believed to be invariant when the process is shifted in time.

1. Squared exponential (SE) kernel:

$$K(t, t') = c^2 \exp\left(-\frac{(t - t')^2}{2\kappa^2}\right), \tag{10}$$

   where $c > 0$ and $\kappa > 0$ are two model parameters, known as the scale and the length scale parameters, respectively.

2. Exponential kernel:

$$K(t, t') = c^2 \exp\left(-\frac{|t - t'|}{2\kappa^2}\right), \tag{11}$$

   where $c > 0$ and $\kappa > 0$ are two model parameters that play similar roles as the ones in the SE kernel above.

3. Periodic kernel (MacKay, 1998):

$$K(t, t') = c^2 \exp\left(-\frac{2\sin^2(\pi|t - t'|/p)}{\kappa^2}\right), \tag{12}$$

   where $c > 0$ and $\kappa > 0$ are two model parameters that play similar roles as the ones in the two kernels above and p is known as the period parameter which determines the periodicity of the kernel function.

Mean zero Gaussian processes with different kernel functions have different properties. For example, the mean zero Gaussian processes with an SE kernel tend to have smooth pathes. In fact, a mean zero Gaussian process with the SE kernel is classified as one of the most smooth stochastic processes, according to the notion of mean square differentiability (Chapter 1, Adler, 1981), a classical quantification of the smoothness of stochastic processes. This kernel function is widely used in statistical applications of Gaussian process. It will be further discussed in the sequel and be used in the data analysis.

An alternative way of parameterizing the kernel is by directly modeling the mean zero Gaussian process, which can be done by using a linear basis function model. Specifically, let $\phi_1(t)$, ..., $\phi_H(t)$ be H pre-specified basis functions on $[0, T]$, such as spline basis, Fourier basis, or wavelet basis functions. The theory of functional principal component analysis provides an idea on choosing better basis functions (e.g., Hall et al., 2008). Given the basis functions, the linear basis function model assumes that

$$\bar{\theta}_i(t) = \sum_{h=1}^{H} \omega_h Z_{ih} \phi_h(t), \tag{13}$$

where ω_h, $h = 1, ..., H$, are model parameters and Z_{ih}, $h = 1, ..., H$, are i.i.d. standard normal random variables. The model (13) yields

$$K(t, t') = \sum_{h=1}^{H} \omega_h^2 \phi_h(t) \phi_h(t').$$

For finite H, this parametrization approach typically leads to a non-stationary kernel function. Making use of the theory of reproducing kernel Hilbert space, essentially any mean zero Gaussian process can be approximated by the form of (13) for sufficiently large H.

Squared exponential kernel. We further discuss on the properties of the SE kernel. According to (10), $Var(\theta_i(t)) = K(t,t) = c^2$. *The scale parameter c thus captures the overall variation of the Gaussian process in the long run.* Moreover, *the length-scale parameter κ captures the short-term temporal dependence.* More precisely, the correlation between $\theta_i(t)$ and $\theta_i(t')$ is given by

$$Cor(\theta_i(t), \theta_i(t')) = \frac{Cov(\theta_i(t), \theta_i(t'))}{\sqrt{(Var(\theta_i(t)) \times Var(\theta_i(t')))}} = \exp\left(-\frac{(t-t')^2}{2\kappa^2}\right).$$

As shown in Figure 4, for each value of κ, the correlation decays towards zero as the time lag increases. The decaying rate is determined by the value of κ. In particular, when the time lag $|t - t'| > 2\kappa$, the correlation is smaller than $\exp(-2) = 0.14$. Moreover, for a given time lag, a smaller value of κ implies a smaller correlation. Figure 5 shows sample paths from three Gaussian processes with mean zero and SE kernels. Specifically, in panel (a), $c = 1, \kappa = 0.5$, in (b), $c = 1, \kappa = 2$, and in (c), $c = 2, \kappa = 2$. Panels (a) and (b) only differ by the values of the κ parameter and the paths in panel (a) are from a Gaussian process with a smaller value of κ. The paths in panel (a) are more wiggly (i.e., have more short-term variation) than those in panel (b), since the Gaussian process in panel (a) has less temporal dependence. Panels (b) and (c) only differ by the values of c, due to which the paths in panel (c) have more variation in the long run.

Identifiability of the model parameters Like many other structural equation models, constraints are needed to ensure model identifiability. In particular, two constraints are needed, one to fix the scale of the latent process and the other to avoid mean shift. For instance, we consider a model combining the mean function (9), the measurement model (4), and the SE kernel (10). To fix the scale in this model, we can either fix the scale parameter $c = 1$ in (10) or the first loading parameter $a_1 = 1$ in (3). In addition, to avoid mean shift, we can set either $\alpha_0 = 0$ in (9) or $b_1 = 0$ in (4).

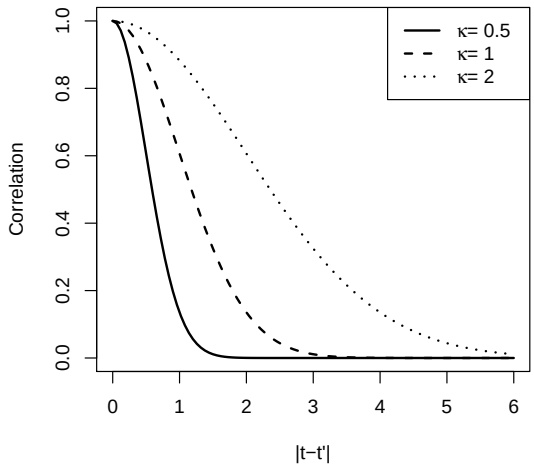


Figure 4: An illustration of the squared exponential kernel.

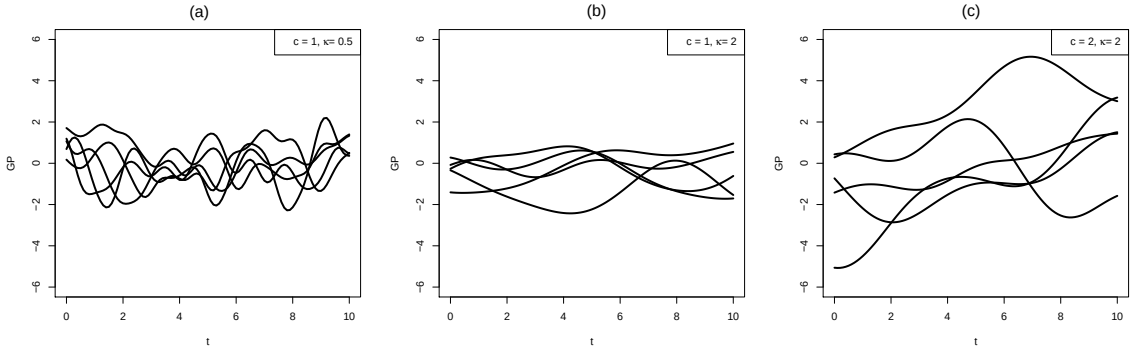


Figure 5: Sample paths from three Gaussian processes with mean 0 and SE kernels. The SE kernels differ by their values of the c and the κ parameters.

3 Inference under LGP Model

The statistical inference under the proposed model can be classified into two levels, the population level and individual level. Both levels of inference may be of interest in the latent curve analysis. The population level inference considers the estimation of the parameters in both the measurement and structural models. The individual level inference focuses on the posterior distribution of $\theta_i(t)$ given data from each individual i when the measurement and the structural models are known (e.g. obtained from the population level inference).

Population level inference. We use Ψ to denote all the model parameters, including parameters from both the measurement and structural models. As mentioned above, constraints may be imposed on Ψ to ensure model identifiability. Our likelihood function can be written as

$$L(\Psi) = \prod_{i=1}^{N} \int \prod_{s=1}^{S_i} \prod_{j=1}^{J} g_j(y_{ij}(t_{is})|\theta_{is}) f_i(\theta_{i1}, ..., \theta_{iS_i}) d\theta_{i1} ... d\theta_{iS_i}, \tag{14}$$

where $f_i(\theta_{i1}, ..., \theta_{iS_i})$ is the density function of an S_i-variate normal distribution with mean $(m(t_{i1}), \cdots m(t_{iS_i}))$ and covariance matrix $(K(t, t') : t, t' = t_{i1}, ..., t_{iS_i})$. Note that this likelihood function is the marginal likelihood of data in which the latent curves are integrated out. The maximum likelihood estimator of Ψ is defined as $\hat{\Psi} = \arg\max_{\Psi} L(\Psi)$, whose computation is discussed in Section 4. We then obtain the estimated mean and kernel functions by plugging in $\hat{\Psi}$.

Individual level inference. Similar to the classical latent curve analysis, the current modeling framework also allows for statistical inference on the latent curve of each individual. For ease of exposition, we assume both the measurement and the structural

models are known when making individual level inference. In practice, we can first estimate the model parameters and then treat the estimated model as the true one in making the individual level inference. For individual i, whether or not measurement occurs at time t^*, one can infer on $\theta_i(t^*)$ based on the posterior distribution of $\theta_i(t^*)$ given $\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})$. By sweeping t^* over the entire interval $[0, T]$, one obtains the posterior mean of $\theta_i(t)$ as a function of t, which serves as a point estimate of individual i's latent curve. When calculated under the estimated model, we call the posterior mean of $\theta_i(t)$ the Expected A Posteriori (EAP) estimate of individual i's latent curve and denote it by $\hat{\theta}_i(t)$. It mimics the EAP estimate of an individual's latent trait level in item response theory (e.g. Embretson & Reise, 2000).

4 Computation

In this section, we elaborate on the computational details.

4.1 Individual Level Inference

We first discuss computing the posterior distribution of $\theta_i(t^*)$ given $\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})$, for any time t^*, when both the measurement and the structural models are given. We denote the density of this posterior distribution by $h(\theta|\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}))$. Following equation (1) of the measurement model, $\theta_i(t^*)$ and $(\mathbf{Y}_i(t_{i1}), ..., \mathbf{Y}_i(t_{iS_i}))$ are conditionally independent given $(\theta_i(t_{i1}), ..., \theta_i(t_{iS_i}))$. Consequently,

$$\begin{aligned} & h(\theta|\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})) \\ = & \int h_1(\theta|\theta_1, ..., \theta_{S_i}) h_2(\theta_1, ..., \theta_{S_i}|\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i})) d\theta_1 ... d\theta_{S_i}, \end{aligned} \tag{15}$$

where $h_1(\theta|\theta_1, ..., \theta_{S_i})$ denotes the conditional distribution of $\theta_i(t^*)$ given $(\theta_i(t_{i1}), ..., \theta_i(t_{iS_i}))$ and $h_2(\theta_1, ..., \theta_{S_i}|\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}))$ denotes the posterior distribution of $(\theta_i(t_{i1}), ..., \theta_i(t_{iS_i}))$ given the observed responses. Specifically, since $(\theta_i(t^*), \theta_i(t_{i1}), ..., \theta_i(t_{iS_i})$ follows a multivariate normal distribution with mean $(m(t^*), m(t_{i1}), ..., m(t_{iS_i}))$ and covariance matrix $(K(t,t') : t, t' = t^*, t_{i1}, ..., t_{iS_i})$, $h_1(\theta|\theta_1, ..., \theta_{S_i})$ is still normal, for which the mean $\mu(\theta_1, ..., \theta_{S_i})$ and variance $\sigma^2(\theta_1, ..., \theta_{S_i})$ have analytic forms. Specifically,

$$\mu(\theta_1, ..., \theta_{S_i}) = m(t^*) + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}) \text{ and } \sigma^2(\theta_1, ..., \theta_{S_i}) = K(t^*, t^*) - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_{S_i})^\top$, $\boldsymbol{\mu} = (m(t_{i1}), ..., m(t_{iS_i}))^\top$, $\Sigma_{12} = (K(t^*, t_{i1}), ..., K(t^*, t_{iS_i}))$, $\Sigma_{22} = (K(t,t') : t, t' = t_{i1}, ..., t_{iS_i})$, and $\Sigma_{21} = \Sigma_{12}^\top$. Then the posterior mean of $\theta_i(t^*)$ is given by

$$\int \mu(\theta_1, ..., \theta_{S_i}) h_2(\theta_1, ..., \theta_{S_i}|\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}))d\theta_1 ... d\theta_{S_i}. \tag{16}$$

In addition, the α-level quantile of the posterior distribution is given by

$$\int (\mu(\theta_1, ..., \theta_{S_i}) + z_\alpha \sigma(\theta_1, ..., \theta_{S_i})) h_2(\theta_1, ..., \theta_{S_i}|\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}))d\theta_1 ... d\theta_{S_i}, \tag{17}$$

where z_α is the α-level quantile of a standard normal distribution.

Under the linear factor model (6), $(\theta_i(t^*), \theta_i(t_{i1}), ..., \theta_i(t_{iS_i}), \mathbf{Y}_i(t_{i1}), ..., \mathbf{Y}_i(t_{iS_i}))$ are jointly normal. Consequently, (15)-(17) have analytical forms. Under other measurement models, (16) and (17) can be approximated by using Monte Carlo samples from the posterior distribution $h_2(\theta_1, ..., \theta_{S_i}|\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}))$. Specifically, let $(\theta_1^{(l)}, ..., \theta_{S_i}^{(l)}), l = 1, ..., L$, be L Monte Carlo samples. Then we approximate the mean and α-level quantile

of the posterior distribution of $\theta_i(t^*)$ by

$$\begin{aligned} &\frac{1}{L}\sum_{l=1}^{L}\mu(\theta_1^{(l)},...,\theta_{S_i}^{(l)}), \\ &\frac{1}{L}\sum_{l=1}^{L}\mu(\theta_1^{(l)},...,\theta_{S_i}^{(l)}) + z_\alpha\sigma(\theta_1^{(l)},...,\theta_{S_i}^{(l)}). \end{aligned} \tag{18}$$

Markov chain Monte Carlo (MCMC) methods can be used to obain Monte Carlo samples from the posterior distribution $h_2(\theta_1,...,\theta_{S_i}|\mathbf{y}_i(t_{i1}),...,\mathbf{y}_i(t_{iS_i}))$. For example, a Gibbs sampler is developed that efficiently samples from this posterior distribution under the probit model (5) for ordinal response data. This sampler, described as follows, makes use of the latent response formulation of the probit model (5).

Step 1: For $i = 1,...,N$, $j = 1,...,J$, $s = 1,...,S_i$, sample $y_{ij}^*(t_{is})$ from a truncated normal distribution that truncates a normal distribution $N(-a_j\tilde{\theta}_i(t_{is}),1)$ by interval $[d_{j,y_{ij}(t_{is})}, d_{j,y_{ij}(t_{is})+1}]$, where $\tilde{\theta}_i(t_{is})$ is some initial value of $\theta_i(t_{is})$.

Step 2: For $i = 1,...,N$,, given $y_{ij}^*(t_{is})$s, we update $(\tilde{\theta}_i(t_{i1}),...,\tilde{\theta}_i(t_{iS_i}))$, by sampling from

$$h_3(\theta_1,...,\theta_{S_i}|\mathbf{y}_i^*(t_{i1}),...,\mathbf{y}_i^*(t_{iS_i})),$$

where $\mathbf{y}_i^*(t) = (y_{i1}^*(t),...,y_{iJ}^*(t))$ and h_3 denotes the conditional distribution of $(\theta_i(t_{i1}),...,\theta_i(t_{iS_i}))$ given the ideal responses $\mathbf{y}_i^*(t_{i1}),...,\mathbf{y}_i^*(t_{iS_i})$. It is worth noting that this conditional distribution is multivariate normal, because $\theta_i(t_{i1})$, ..., $\theta_i(t_{iS_i})$, $\mathbf{y}_i^*(t_{i1})$, ..., $\mathbf{y}_i^*(t_{iS_i})$ are jointly normal. The observed data $\mathbf{y}_i(t_{i1})$, ..., $\mathbf{y}_i(t_{iS_i})$ are not conditioned upon, because $\theta_i(t_{i1})$, ..., $\theta_i(t_{iS_i})$ are conditionally independent of the observed data when given the latent responses $\mathbf{y}_i^*(t_{i1})$, ..., $\mathbf{y}_i^*(t_{iS_i})$.

We point out that both steps can be efficiently computed, because step 1 only involves sampling from univariate truncated normal distributions and step 2 only involves sam-

pling from multivariate normal distributions. Well-developed samplers exist for both steps.

4.2 Population Level Inference

We now discuss the computation for maximizing the likelihood function (14). Under the linear factor model (6), the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) is used to optimize (14), where the E-step is in a closed form due to the joint normality of data and latent variables. The implementation of this EM algorithm is standard and thus we omit the details here.

Under other measurement models, the classical EM algorithm is typically computationally infeasible when the number of time points is large, in which case the E-step of the algorithm involves a high-dimensional integral that does not have an analytical form. We adopt a stochastic EM (StEM) algorithm (Celeux & Diebolt, 1985; Diebolt & Ip, 1996; Zhang et al., 2018) which avoids the numerical integration in the E-step of the standard EM algorithm (Dempster et al., 1977; Bock & Aitkin, 1981) by Monte Carlo simulations. The convergence properties of the StEM algorithm are established in Nielsen (2000). Similar to the EM algorithm, the StEM algorithm iterates between two steps, the StE step and the M step. Let $\Psi^{(0)}$ be the initial parameter values and $(\tilde{\theta}_{i1}^{(0)}, \cdots, \tilde{\theta}_{iS_i}^{(0)}), i = 1, ..., N$, be the initial values of person parameters. In each step l ($l \geq 1$), the following StE step and M step are performed.

StE step: For $i = 1, ..., N$, sample $(\tilde{\theta}_{i1}^{(l)}, \cdots, \tilde{\theta}_{iS_i}^{(l)})$ from

$$h_2(\theta_1, ..., \theta_{S_i} | \mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}); \Psi^{(l-1)}),$$

the conditional distribution of $(\theta_i(t_{i1}), \cdots, \theta_i(t_{iS_i}))$ given $(\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}))$ under parameters $\Psi^{(l-1)}$. For the probit model (5), we use the Gibbs sampler described

in Section 4.1 to sample from $h_2(\theta_1, ..., \theta_{S_i} | \mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}); \Psi^{(l-1)})$.

M step: Obtain parameter estimate

$$\Psi^{(l)} = \arg\max_{\Psi} \sum_{i=1}^{N} l(\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}), \tilde{\theta}_{i1}^{(l)}, \cdots, \tilde{\theta}_{iS_i}^{(l)}; \Psi), \tag{19}$$

where

$$\begin{aligned} & l(\mathbf{y}_i(t_{i1}), ..., \mathbf{y}_i(t_{iS_i}), \tilde{\theta}_{i1}^{(l)}, \cdots, \tilde{\theta}_{iS_i}^{(l)}; \Psi) \\ = & \sum_{s=1}^{S_i} \left[\sum_{j=1}^{J} \log g_j(y_{ij}(t_{is}) | \tilde{\theta}_{is}^{(l)}) \right] + \log f_i(\tilde{\theta}_{i1}^{(l)}, ..., \tilde{\theta}_{iS_i}^{(l)}) \end{aligned} \tag{20}$$

is the complete data log-likelihood of a single observation. Note that g_j and f_i are defined in (3) and (14), respectively, containing model parameters. In our implementation, the optimization is done using the L-BFGS-B algorithm (Liu & Nocedal, 1989).

The final estimate of Ψ is given by the average of $\Psi^{(l)}$s from the last m iterations, i.e.,

$$\hat{\Psi} = \frac{1}{m} \sum_{l=m_0+1}^{m_0+m} \Psi^{(l)}. \tag{21}$$

As shown in Nielsen (2000), $\hat{\Psi}$ can approximate the maximum likelihood estimator sufficiently accurately, when m_0 and m are large enough.

5 Incorporation of Covariates

In practice, individual specific covariates are often collected and incorporated into the latent curve analysis. As visualized in the path diagram in Figure 6, covariates $\mathbf{x}_i$ can be further added to the structural model to explain how the distribution of the latent curves depends on the covariates. A specific type of covariates of interest is group membership,

such as experimental versus control and female versus male. Latent curve analysis that incorporates discrete group membership as covariates in the structural model is referred to as the analysis of groups (Chapter 6, Bollen & Curran, 2006).

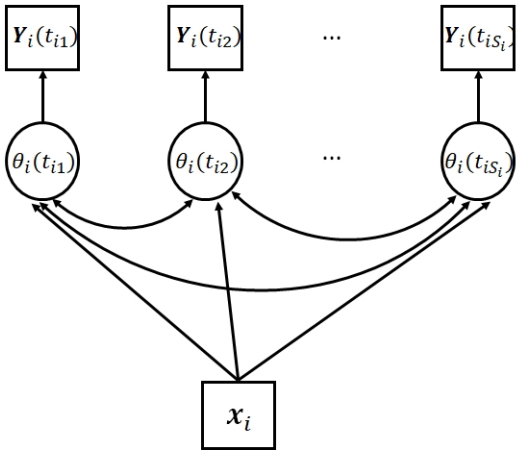


Figure 6: Path diagram of a latent curve model with covariates.

Covariates can be easily handled under the proposed framework. For example, when discrete group membership may affect the mean function of the latent curve, we let parameters in $m(t)$ to be group-specific. Similarly, we may also allow parameters in $K(t, t')$ to depend on the group membership. Quantitative covariates, such as age, can also be incorporated into the current model. The mean and kernel functions are denoted by $m_{\mathbf{x}_i}(t)$ and $K_{\mathbf{x}_i}(t, t')$ when they depend on the covariates. The tools for the inference and computation discussed above can be easily generalized.

6 Simulation

The proposed modeling framework and the estimation procedures are further evaluated by simulation studies.

6.1 Study I

We first evaluate the parameter recovery using the EM algorithm, under a setting similar to the real data example in Section 7, except that a single group is considered in this study. In particular, it is assumed that each participant is measured for 25 consecutive days, with four measurements per day. Such a design results in 100 times of measurement. The time points of the four measurements are randomly sampled within a day. And we consider a measurement model with a single indicator. More precisely, given the observation time, the model is specified as follows.

$$Y_{i1}(t)|\theta_i(t) \sim N(\theta_i(t), \sigma^2),$$
$$\theta_i(\cdot) \sim GP(m, K),$$

where $m(t) = \alpha$ and $K(t, t') = c^2 \exp(-(t - t')^2/(2\kappa^2))$. The true model parameters are specified in Table 1. Two sample sizes are considered, including $N = 50$ and $N = 100$. The simulation under each sample size is repeated for 100 times, based on which the mean squared error (MSE) for parameter estimation is calculated. According to the MSE for parameter estimation presented Table 1, the parameter estimation is very accurate under the current simulation settings and the estimation accuracy improves as the sample size increases.

We further illustrate the performance of the individual level inference based on the L^2 distance between $\theta_i(t)$ and its EAP estimate $\hat{\theta}_i(t)$, where the distance is defined as

$$d_i = \sqrt{\int_0^T (\theta_i(t) - \hat{\theta}_i(t))^2 dt}.$$

In particular, d_i quantifies the inaccuracy of estimating the latent curve $\theta_i(t)$ by $\hat{\theta}_i(t)$.

	α	c^2	κ	σ^2
True	1.5	0.4	0.3	0.1
MSE($N = 50$)	2.7×10^{-4}	2.2×10^{-4}	1.7×10^{-4}	1.3×10^{-5}
MSE($N = 100$)	1.2×10^{-4}	9.2×10^{-5}	1.5×10^{-4}	1.2×10^{-5}

Table 1: Simulation Study I: Simulation results on the parameter recovery for an LGP model with a linear factor measurement model.

The L^2 distance between $\theta_i(t)$ and $\hat{\alpha}$,

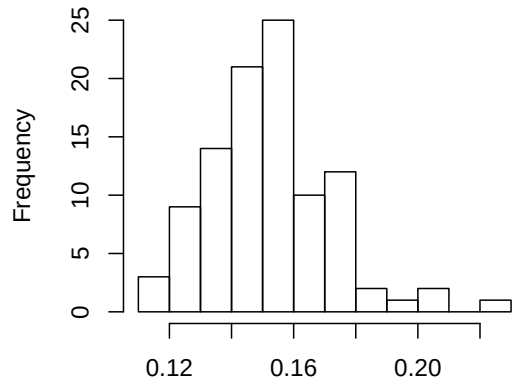
$$e_i = \sqrt{\int_0^T (\theta_i(t) - \hat{\alpha})^2 dt},$$

is used as a reference for d_i that quantifies the inaccuracy of estimating $\theta_i(t)$ by the estimate of the population mean $\hat{\alpha}$. The ratio d_i/e_i serves as a measure of inaccuracy in estimating the latent curve of individual i, in which the difficulty in estimating the curve has been taken into account by the denominator e_i. The smaller the ratio is, the more accurate the latent curve $\theta_i(t)$ is estimated in a relative sense (relative to the overall difficulty in estimating $\theta_i(t)$ measured by e_i).

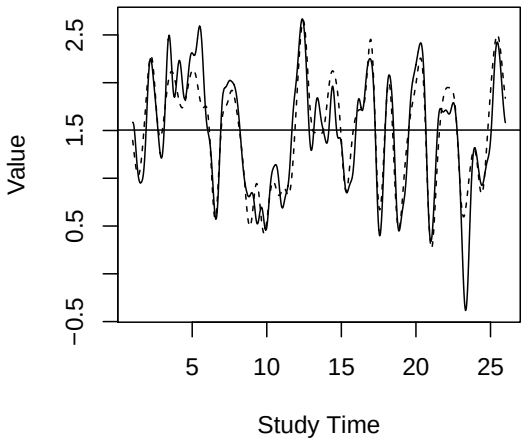
In panel (a) of Figure 7, we show the histogram of the ratios d_i/e_i for all individuals from a randomly selected dataset among all replications when the sample size $N = 100$. As we can see, d_i is much smaller than e_i, implying that $\hat{\theta}_i(t)$ estimates $\theta_i(t)$ very accurately. Panels (b)-(d) of Figure 7 show $\theta_i(t)$, $\hat{\theta}_i(t)$, as well as $\hat{\alpha}$ for three randomly selected individuals from the same dataset. According to these plots, the true latent curves are well approximated by their EAP estimates.

6.2 Study II

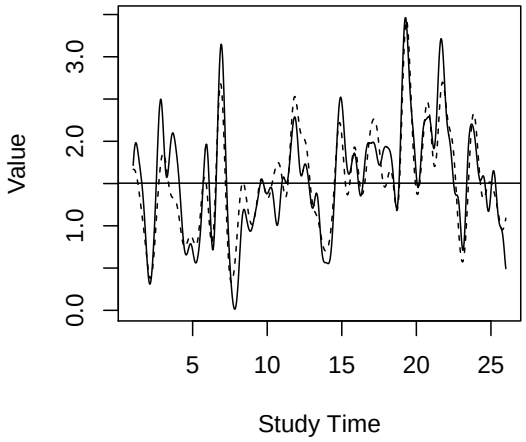
We now consider a simulation study whose setting is the same as Study I except for a different measurement model component. In particular, we consider ordinal response



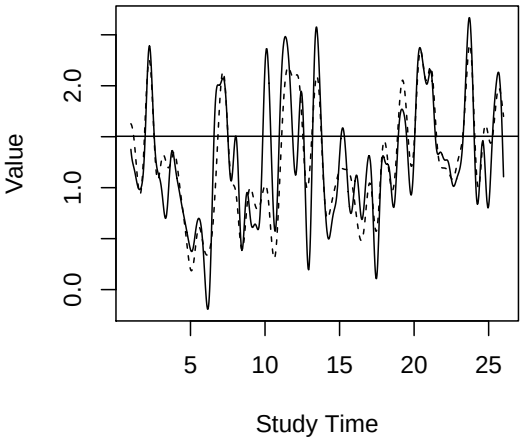
(a) Histogram of the ratios d_i/e_i for all individuals from a randomly selected dataset.



(b) $\theta_i(t)$ (solid line) versus $\hat{\theta}_i(t)$ for an individual with $d_i/e_i = 0.14$



(c) $\theta_i(t)$ (solid line) versus $\hat{\theta}_i(t)$ for an individual with $d_i/e_i = 0.15$



(d) $\theta_i(t)$ (solid line) versus $\hat{\theta}_i(t)$ for an individual with $d_i/e_i = 0.20$

Figure 7: Simulation Study I: Results on individual level inference.

data generated by the probit model (5). Specifically, the measurement at each time point is assumed to be based on five polytomous items, each with three ordinal categories (i.e., $n_j = 3$). The true model parameters are given in Table 2. Note that we fix $a_1 = 1$ and $d_{1,1} = 0$ in both the true model and the estimation procedure for model identifiability.

The simulation under each sample size is repeated for 100 times. For each simulated dataset, the model parameters are estimated using the stochastic EM algorithm described in Section 4.2, based on a random initial value. The two tuning parameters m_0 and m of the algorithm are set to be 100 and 200, respectively. The estimation accuracy measured by mean squared error is shown in Table 2, which indicates an accurate estimation result. The running time of the stochastic EM algorithm for one dataset with $N = 100$ is around 10 minutes[1]. It can be further speeded up by parallel computing.

Finally, we examine the recovery of the individual latent curves, measured by the L_2 distance ratio d_i/e_i defined in Study I. The EAP estimates of the individual curves are obtained by Monte Carlo approximation (18), where $L = 100$ Monte Carlo samples are used. In particular, the histogram of d_i/e_i, $i = 1, ..., N$, is presented in Figure 8, for a randomly selected dataset among all replications under $N = 100$. According to the histogram, d_i is much smaller than e_i, though the ratios tend to be larger than those in Study I. It implies that, under the current setting, the EAP estimate $\hat{\theta}_i(t)$ is still substantially more accurate than the population mean $\hat{\alpha}$ in estimating all individuals' latent curves.

[1]The study is conducted on a personal computer with specifications: Processor 2.2 GHz Intel Core i7; Memory 8 GB 1600 MHz DDR3.

	a_1	a_2	a_3	a_4	a_5
True	1.00	1.00	0.65	0.62	0.53
MSE(N=50)	·	1.7×10^{-3}	7.9×10^{-4}	5.6×10^{-4}	4.1×10^{-4}
MSE(N=100)	·	7.5×10^{-4}	3.2×10^{-4}	2.7×10^{-4}	2.1×10^{-4}
	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	$d_{4,1}$	$d_{5,1}$
True	0.00	0.45	-0.25	-0.27	0.34
MSE(N=50)	·	1.4×10^{-3}	6.2×10^{-4}	7.6×10^{-4}	6.5×10^{-4}
MSE(N=100)	·	6.9×10^{-4}	3.3×10^{-4}	3.5×10^{-4}	3.8×10^{-4}
	$d_{1,2}$	$d_{2,2}$	$d_{3,2}$	$d_{4,2}$	$d_{5,2}$
True	1.84	1.45	0.44	1.37	1.55
MSE(N=50)	1.7×10^{-3}	2.1×10^{-3}	6.9×10^{-4}	1.0×10^{-3}	1.1×10^{-3}
MSE(N=100)	6.7×10^{-4}	1.2×10^{-3}	4.3×10^{-4}	5.5×10^{-4}	5.4×10^{-4}
	α	κ	c^2		
True	-0.79	0.30	1.27		
MSE(N=50)	1.6×10^{-3}	2.9×10^{-4}	2.0×10^{-3}		
MSE(N=100)	1.1×10^{-3}	3.0×10^{-4}	9.3×10^{-4}		

Table 2: Simulation Study II: Simulation results on the parameter recovery accuracy for an LGP model with a probit measurement model component.

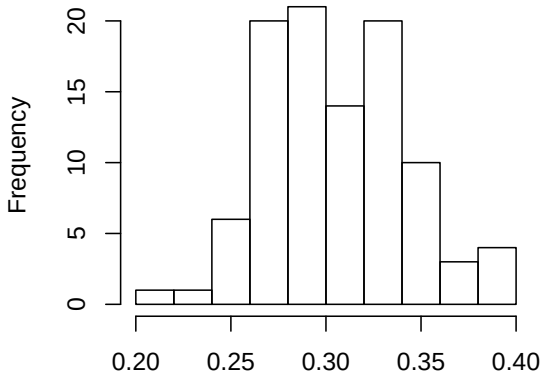


Figure 8: Simulation Study II: Histogram of the ratios d_i/e_i for all individuals from a randomly selected dataset under $N = 100$.

7 Analysis of Negative Mood in BPD and MDD/DYS Patients

We analyze data from a study of the affective instability in borderline personality disorder (Trull et al., 2008) that collected ecological momentary assessment data from psychiatric outpatients with borderline personality disorder (BPD) and with major depressive disorder (MDD) or dysthymic disorder (DYS). The participants were recruited from one of four community mental health outpatient clinics through flyers. The dataset has been analyzed in Jahng et al. (2008) and is downloaded from `http://dx.doi.org/10.1037/a0014173.supp`. The data contain 84 participants: 46 who met DSM-IV-TR (American Psychiatric Association, 2000) diagnostic criteria for BPD and who endorsed the diagnostic feature of affective instability; and 38 who met DSM-IV-TR diagnostic criteria for current MDD or DYS and did not report affective instability.

This dataset contains, for each time and each participant, a negative affect composite score based on 21 items from the Positive and Negative Affect Scales-Extended Version (Watson & Clark, 1999). The participants were measured multiple times a day over approximately 4 weeks of consecutive days. As commonly encountered in EMA data, the number of days of assessments per person and the number of assessments per day differed (days per person: median = 29, interquartile range = 2; assessments per day: median = 5, interquartile range = 1). In total, the participants received 76 to 186 assessments (median = 153, interquartile range = 24) per person were conducted. Table 3 illustrates the data structure, where the five columns show the individual ID, the negative affect composite score, the group membership ($x_i = 0$ for the MDD/DYS group, $x_i = 1$ for the BPD group), the study time, and the calendar time, respectively. In particular, the study time uses day as the time unit and sets 00:00 of the first day receiving measurement as time 0 for each individual. Figure 9 visualizes the data from a

ID	Score	Group	Study Time	Calendar Time
1	1.19	0	0.74	2005-03-18 17:40:00
1	1.81	0	1.52	2005-03-19 12:24:38
1	1.38	0	1.63	2005-03-19 15:06:36
1	1.86	0	1.66	2005-03-19 15:49:34
⋮	⋮	⋮	⋮	⋮

Table 3: An illustration of the EMA data from the mood study of BPD and MDD/DYS patients.

MDD/DYS patient and that from a BPD patient, where the individuals receive different numbers of measurement, at different and unequally spaced time points.

Following the research question of Jahng et al. (2008), we investigate, by making use of the proposed latent Gaussian process model, whether the BPD group suffers from more temporal negative mood instability than the MDD/DYS group. We also investigate the mean of the negative mood of the two groups. To answer these questions under the latent Gaussian process modeling framework, we treat the negative affect composite score as a continuous variable and adopt a single-indicator linear factor measurement model. In addition, we assume the mean and the kernel functions of the latent Gaussian process are group specific. Specifically, the model is specified as follows.

$$Y_{i1}(t)|\theta_i(t) \sim N(\theta_i(t), \sigma^2),$$
$$\theta_i(\cdot)|x_i \sim GP(m_{x_i}, K_{x_i}),$$

where $m_0(t) = \alpha_0$, $m_1(t) = \alpha_1$, $K_0(t,t') = c_0^2 \exp\left(-(t-t')^2/(2\kappa_0^2)\right)$, and $K_1(t,t') = c_1^2 \exp\left(-(t-t')^2/(2\kappa_1^2)\right)$. Under these assumptions, the Gaussian process for each group is stationary. According to the recruitment design of the study, the stationarity assumption seems reasonable.

The main results are shown in Table 4, including parameter estimates obtained from the EM algorithm and their 95% bootstrap confidence interval (Chapter 6, Efron &

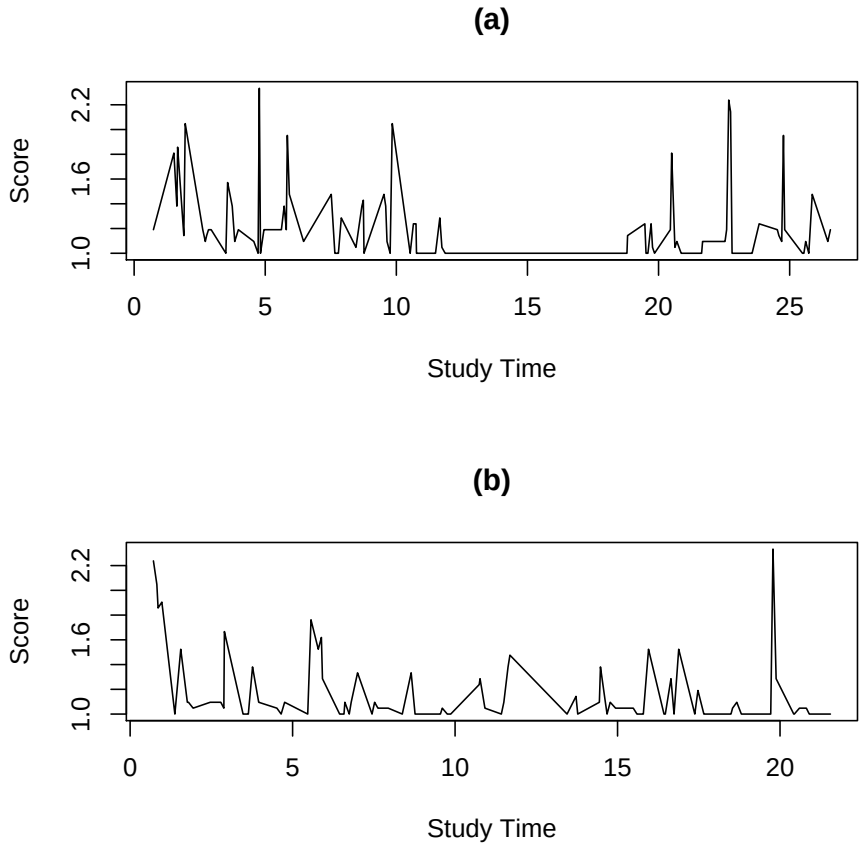


Figure 9: An illustration of the EMA data, where panels (a) and (b) show the negative affect composite score (y-axis) versus the study time (x-axis) from a MDD/DYS patient and a BPD patient, respectively.

Tibshirani, 1993). The bootstrap results are obtained by resampling individuals with replacement. In particular, an estimate of the variance due to the measurement error is $\hat{\sigma}^2 = 0.091$, which is much smaller than $\hat{c}_0^2 = 0.234$ and $\hat{c}_1^2 = 0.440$, the overall variations of the two Gaussian processes. In addition, the two groups only significantly differ by the overall long-run variations, with a difference $\hat{c}_1^2 - \hat{c}_0^2 = 0.206$ which has a corresponding 95% bootstrap confidence interval $(0.010, 0.407)$. That is, the BPD group has more variation in the long run than the MDD/DYS group, which is consistent with the existing knowledge these mental health disorders. Their overall mean scores are not significantly different, for which the difference is $\hat{\alpha}_1 - \hat{\alpha}_0 = 0.081$ and a 95% confidence interval $(-0.106, 0.275)$. Similarly, the two groups do not significantly differ in terms of the short-term temporal dependence, evidenced by $\hat{\kappa}_1 - \hat{\kappa}_0 = 0.012$ and its 95% confidence interval $(-0.039, 0.060)$.

In addition to the estimation of the model parameters, the proposed modeling framework allows us to make inference at the individual level. To demonstrate, in Figure 10, we show the posterior mean and the posterior 2.5% and 97.5% quantiles of $\theta_i(t)$, as well as the corresponding response process, of four participants, two of whom are from the MDD/DYS group and the other two from the BPD group. The calculation of the posterior mean and the posterior quantile for $\theta_i(t)$ is described in Section 4. As we can see, the posterior mean of $\theta_i(t)$ is quite smooth and captures the overall trend of the response process. In addition, the confidence band, given by the posterior 2.5% and 97.5% quantiles of $\theta_i(t)$, becomes wide when two subsequent measurements have a long time lag. For example, participant 35 from the BPD group did not have measurement from the 11th to the 13th day and from the 22nd to the 27th day. That is why the wide confidence bands are observed in panel (d) within the corresponding intervals. When there are multiple measurements occur around a single time point t, the posterior variance at time t can be close to 0 and consequently the corresponding posterior mean and

	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{c}_0^2$	$\hat{c}_1^2$	$\hat{\kappa}_0$	$\hat{\kappa}_1$	$\hat{\sigma}^2$
Point estimate	1.549	1.630	0.234	0.440	0.237	0.249	0.091
95% CI lower bound	1.436	1.476	0.154	0.273	0.192	0.201	0.071
95% CI upper bound	1.658	1.793	0.304	0.608	0.272	0.281	0.112

Table 4: Results from fitting an LGP model to the EMA data from a mood study of BPD and MDD/DYS patients.

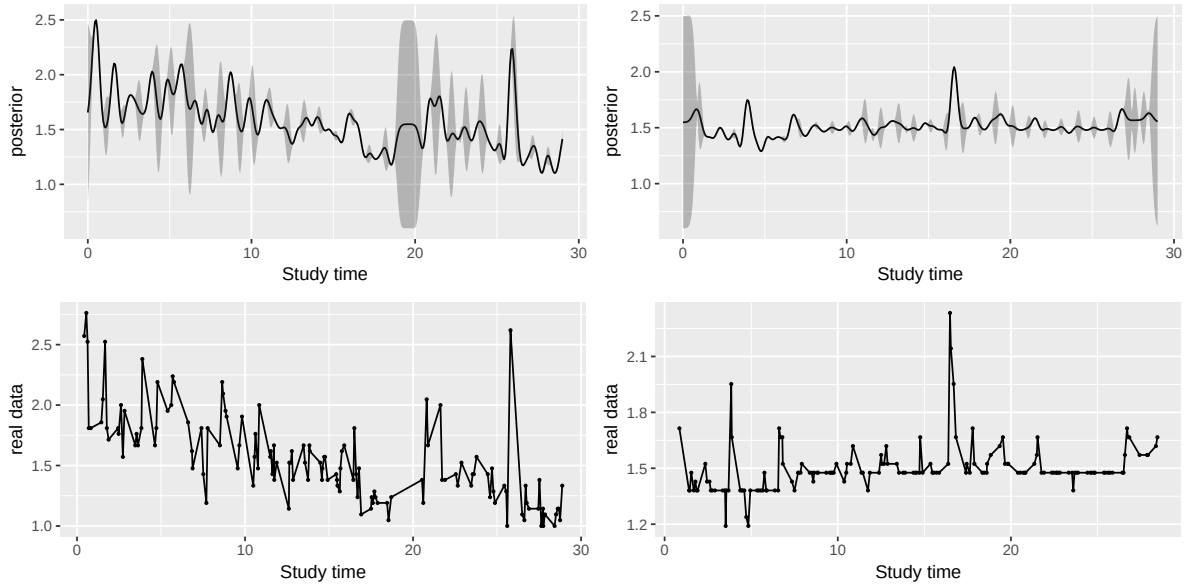
posterior 2.5% and 97.5% quantiles are close to each other.

8 Concluding Remarks

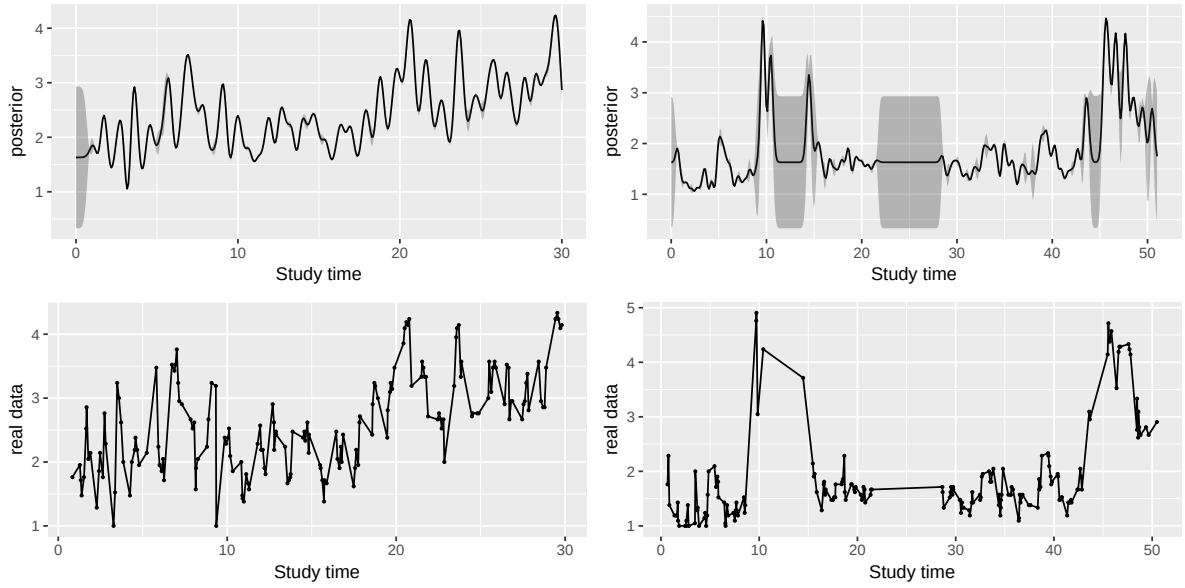
In this paper, we introduce the latent Gaussian process model as a general family of continuous-time latent curve models. This new model complements the existing models for the analysis of intensive longitudinal data. The proposed model decomposes the latent curve analysis into a measurement model component and a structural model component. The measurement component captures the conditional distribution of an individual's observed data given his/her latent curve in a continuous time domain and the structural component models the distribution of the latent curve. It is shown that many existing latent curve models are special cases of the proposed one.

In particular, a Gaussian process model is proposed for the modeling of latent curves in the structural model component. By making use of the mathematical properties of Gaussian processes, the modeling of the structural component is further decomposed into separate modeling of the mean function and the Kernel function of a Gaussian process. Estimation and statistical inference are further discussed under an empirical Bayes framework, where inference is considered at both population and individual levels.

The proposed model and methods are further illustrated through simulation studies and a real data example. In particular, our analysis of the negative mood of BPD and MDD/DYS patients reveals that the main difference between the two groups is due to



(a) Participant 47 (from the MDD/DYS group)(b) Participant 17 (from the MDD/DYS group)



(c) Participant 9 (from the BPD group) (d) Participant 35 (from the BPD group)

Figure 10: The posterior mean and the posterior 2.5% and 97.5% quantiles of $\theta_i(t)$, as well as the corresponding response process, of four participants. Participants in panels (a) and (b) are from the MDD/DYS group and participants in panels (c) and (d) are from the BPD group.

the BPD group having significantly higher long-term variation, while the two groups are not significantly different in the mean negative affect levels and in the short-term temporal dependence.

The proposed framework leads to many new directions, which are left for future investigation. First, it is often of interest to measure multiple correlated dynamic latent traits, in which case $\theta_i(t)$ becomes a vector at each time point t. The current framework can be easily extended to that setting, by adopting a multidimensional measurement model (e.g., multidimensional item response theory model) and a multivariate Gaussian process model for the structural component. Second, many intensive longitudinal studies involve not only measurement but also interventions (e.g., treatment of mental health disorders). Interventions can be viewed as time-dependent covariates which can be incorporated into the structural component of the proposed model. By estimating the coefficients associated with the intervention covariates, the intervention effects can be evaluated dynamically. Finally, the psychometric properties of the proposed model remain to be studied, such as the detection of differential item functioning, the assessment of model goodness-of-fit, and the evaluation of measurement reliability.

References

Adler, R. J. (1981). *The geometry of random fields*. New York, NY: John Wiley & Sons.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders 4th ed., revised*. Washington, DC: American Psychiatric Association.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 359–388. doi: 10.1080/10705511.2017.1406803

Bianconcini, S., & Bollen, K. A. (2018). The latent variable-autoregressive latent trajectory model: A general framework for longitudinal data analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 791–808. doi: 10.1080/10705511.2018.1426467

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459. doi: 10.1007/bf02294168

Bolger, N., & Laurenceau, J. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York, NY: Guilford Press.

Bollen, K., & Curran, P. (2006). *Latent curve models: A structural equation perspective.* New York, NY: John Wiley & Sons.

Celeux, G., & Diebolt, J. (1985). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, *2*, 73–82.

Conner, T. S., & Lehman, B. (2012). *Handbook of research methods for studying daily life.* New York, NY: Guilford Press.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*, 1–38. doi: 10.1142/9789812388759_0028

Diebolt, J., & Ip, E. H. (1996). Stochastic EM: Method and application. In W. R. Gilks, S. Richardson, & D. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 259–273). New York, NY: CRC Press.

Duncan, T., Duncan, S., & Strycker, L. (2013). *An introduction to latent variable growth curve modeling: Concepts, issues, and application*. New York, NY: Taylor & Francis.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Hall, P., Müller, H.-G., & Yao, F. (2008). Modelling sparse generalized longitudinal observations with latent gaussian processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*, 703–723. doi: doi.org/10.1111/j.1467-9868.2008.00656.x

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine*, *31*, 3328–3336. doi: 10.1002/sim.5338

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, *13*, 354–375. doi: 10.1037/a0014173

Liu, D. C., & Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, *45*, 503–528. doi: 10.1007/BF01589116

Lu, Z.-H., Chow, S.-M., Sherwood, A., & Zhu, H. (2015). Bayesian analysis of ambulatory blood pressure dynamics with application to irregularly spaced sparse data. *The Annals of Applied Statistics*, *9*, 1601–1620. doi: 10.1214/15-aoas846

MacKay, D. J. (1998). Introduction to Gaussian processes. *NATO ASI Series F Computer and Systems Sciences*, *168*, 133–166.

Nielsen, S. F. (2000). The stochastic EM algorithm: Estimation and asymptotic results. *Bernoulli*, *6*, 457–489. doi: 10.2307/3318671

Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, *65*, 199–215. doi: 10.1007/bf02294374

Ram, N., & Grimm, K. J. (2015). Growth curve modeling and longitudinal factor analysis. In R. M. Lerner, W. F. Overton, & P. C. M. Molenaar (Eds.), *Handbook of child psychology and developmental science, volume 1* (p. 758-788). New York, NY: Wiley. doi: 10.1002/9781118963418.childpsy120

Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis*. New York, NY: Springer.

Rasmussen, C. E., & Williams, C. K. (2005). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. New York, NY: CRC Press.

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*, *117*, 647–661. doi: 10.1037/a0012532

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review*, *36*, 823–841. doi: 10.1103/physrev.36.823

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*, 176–192. doi: 10.1037/a0027543

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form.* Ames, IA: The University of Iowa.

Zhang, S., Chen, Y., & Liu, Y. (2018). An improved stochastic EM algorithm for large-scale full-information item factor analysis. *British Journal of Mathematical and Statistical Psychology*. (In press) doi: doi.org/10.1111/bmsp.12153