

FITTING NONLINEAR ORDINARY DIFFERENTIAL EQUATION MODELS WITH
RANDOM EFFECTS AND UNKNOWN INITIAL CONDITIONS USING THE
STOCHASTIC APPROXIMATION EXPECTATION–MAXIMIZATION (SAEM)
ALGORITHM

SY- MIIN CHOW

THE PENNSYLVANIA STATE UNIVERSITY

ZHAOHUA LU AND HONGTU ZHU

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

ANDREW SHERWOOD

DUKE UNIVERSITY

The past decade has evidenced the increased prevalence of irregularly spaced longitudinal data in social sciences. Clearly lacking, however, are modeling tools that allow researchers to fit dynamic models to irregularly spaced data, particularly data that show nonlinearity and heterogeneity in dynamical structures. We consider the issue of fitting multivariate nonlinear differential equation models with random effects and unknown initial conditions to irregularly spaced data. A stochastic approximation expectation–maximization algorithm is proposed and its performance is evaluated using a benchmark nonlinear dynamical systems model, namely, the Van der Pol oscillator equations. The empirical utility of the proposed technique is illustrated using a set of 24-h ambulatory cardiovascular data from 168 men and women. Pertinent methodological challenges and unresolved issues are discussed.

Key words: differential equation, dynamic, nonlinear, stochastic EM, longitudinal.

From difference scores (see e.g., Bereiter, 1963; Cronbach & Furby, 1970; Harris, 1963) to confirmatory models grounded on differential/difference equations, the study of change remains a central question of interest to researchers in the social and behavioral sciences. In the realm of nonlinear dynamic systems analysis, the last decade has evidenced a gradual shift from heavy reliance on geometrically based *exploratory* nonlinear analytic techniques (see e.g., Kaplan & Glass, 1995; Longstaff & Heath, 1999) to *confirmatory* approaches of studying nonlinear dynamic processes via model fitting (Molenaar & Newell, 2003; Ramsay, Hooker, Campbell, & Cao, 2007).

Differential equation models provide a direct representation of change processes while allowing the data to be *irregularly spaced*. Such data have become increasingly prevalent in studies aimed at collecting experience sampling or ecological momentary assessment data (Stone & Shiffman, 1994). Most experience sampling studies require respondents to provide assessments on relevant constructs over a specified period at specific times of the day (interval-contingent), at random times when prompted by an experimenter-invoked signal (signal-contingent), or as triggered by an event in everyday life (event-contingent; Bolger, Davis, & Rafaeli, 2003). Signal- and event-contingent data are typically irregularly spaced by nature of the study designs.

Currently, there has been a scarcity of tools for fitting models to irregularly spaced data in the psychometric literature. Standard growth curve and the related mixed effects models (Browne & du Toit, 1991; Mcardle & Hamagami, 2003; Meredith & Tisak, 1990) provide a straightforward way of handling irregularly spaced intensive repeated measures data when time appears

Correspondence should be made to Sy-Miin Chow, The Pennsylvania State University, 413 Biobehavioral Health Building, University Park, PA 16802 USA. Email: symiin@psu.edu

explicitly in the fitted functions. When that is not the case, this approach cannot be used without further modifications. Differential equation models, in contrast, can be used to extend conventional growth curve models in a number of ways. First, most growth curve (e.g., linear, Gompertz and exponential, among many others) models can be viewed as the integral solutions of various differential equations. Thus, growth curve models can be conceived as special cases of differential equation models. Second, differential equation models, when compared to standard growth curve expressions, focus explicitly on representing the mechanisms of change: that is, changes in the constructs of interest appear explicitly on the left-hand-side of the equations. In this way, differential equations have greater flexibility in capturing the interdependencies among multiple change processes, especially when the fitted functions do not depend explicitly on time.

Despite the proliferation of work on ODE modeling in the econometric, engineering, and statistical literature (Ait-Sahalia, 2008; Jones, 1984; Mbalawata, Särkkä, & Haario, 2013; Beskos, Papaspiliopoulos, Roberts, & Fearnhead, 2006; Beskos, Papaspiliopoulos, & Roberts, 2009; Ramsay et al., 2007; Särkkä, 2013), much of the progress in ODE modeling in the field of psychometrics has been limited to linear ODE modeling. Notable advances include efforts to extend earlier approaches of fitting the nonlinear integral solutions of linear ordinary differential equations (ODEs) as linear structural equation models without the necessary constraints (Arminger, 1986) to alternative state-space approaches (Jones, 1984, 1993), SEM implementation of state-space approaches (Singer, 1992, 2010, 2012; Oud & Jansen, 2000; Oud & Singer, 2010), two-stage derivative estimation approaches (Boker & Graham, 1998; Boker & Nesselroade, 2002), as well as comparisons between the two-stage and other single-stage approaches (Oud, 2007). Still, these approaches were designed primarily to fit longitudinal linear ODEs and stochastic differential equations (SDEs). Generalizing standard SEM procedures (e.g., product indicator techniques or related approaches involving nonlinear constraints; Kenny & Judd, 1984; Klein & Muthén, 2007; Marsh, Wen, & Hau, 2004) to nonlinear dynamic models (e.g., nonlinear ODEs) is far from simple. Implementing such constraints in the simpler linear growth curve models has proven to be difficult (Duncan, Duncan, Strycker, Li, & Alpert, 1999; Li, Duncan, & Acock, 2000; Wen, Marsh, & Hau, 2002), not to mention other less widely tested nonlinear dynamic models.

Despite the difficulties involved, nonlinear differential equation models have distinct merits compared to linear models that make the associated efforts worth pursuing. For instance, nonlinear differential equation models are capable of predicting ongoing oscillations between different locally stable behaviors (Hale & Koçak, 1991). In this vein, Boker and Graham (1998) used a cubic oscillator model to represent adolescent substance abuse behavior as having two attractor states: substance use and non-use. Perhaps of particular interest to social and behavioral scientists is nonlinear differential equation models' flexibility in representing the dependencies of a change process on—or in other words, how the process is *moderated by*—other key variables in the system. Examples of applications along this line include the use of a modified Van der Pol oscillator equation to represent human circadian rhythms (Brown & Luithardt, 1999; Brown, Luithardt, & Czeisler, 2000), and variations of the predator–prey model to represent dyadic interaction (Chow, Ferrer, & Nesselroade, 2007), human cerebral development (Thatcher, 1998), and cognitive aging (Chow & Nesselroade, 2004). Thus, differential equation models share the merits of their discrete-time counterparts (such as state-space models and the SEM-based latent difference approach; Durbin & Koopman, 2001; McArdle & Hamagami, 2001) in providing a platform to represent change mechanisms in concrete terms, while offering more flexibility in accommodating irregular time intervals.

In the present article, we present a frequentist approach to fitting nonlinear ODE models with random effects and unknown initial conditions by means of a stochastic approximation expectation–maximization (SAEM) algorithm. The proposed approach extends previous work on fitting ODEs in the statistical, biostatistical, and psychometric literature in several ways. First, contrary to other existing linear approaches (Boker & Nesselroade, 2002; Oud & Jansen, 2000; Jones,

1984), we consider the problem of fitting nonlinear ODEs to irregularly spaced data. The proposed approach can also be used with linear ODEs. Second, unlike other applications using a fully Bayesian approach (e.g., Carlin, Gelfand, & Smith, 1992; Chow, Tang, Yuan, Song, & Zhu, 2011; Durbin & Koopman, 2001; Geweke & Tanizaki, 2001; Mbalawata et al., 2013; Särkkä, 2013), we combine a Markov chain Monte Carlo (MCMC) procedure with the expectation–maximization (EM) algorithm to yield maximum likelihood (ML) point and standard error estimates of the time-invariant modeling parameters (as in Kuhn & Lavielle, 2005; Donnet & Samson, 2007). Considerable modeling flexibility is gained due to the ease with which MCMC procedures handle more complex models. Yet, we can still adopt familiar Frequentist-based statistics and approaches (e.g., confidence intervals) for inferential purposes. Third, we represent the structural parameters (i.e., parameters that govern the dynamics of a system) as composed of a series of fixed and random effects—a modeling feature not considered in other studies, including studies that utilize other Frequentist and/or simulation-based approximation approaches to fit ODEs or SDEs (e.g., Beskos et al., 2009, 2006; Chow et al., 2007; Ramsay et al., 2007; Gordon, Salmond, & Smith, 1993; Hürzeler & Künsch, 1998; Kitagawa, 1998; Mbalawata et al., 2013; Singer, 1995, 2002, 2007; Tanizaki, 1996). Finally, while the performance of the SAEM algorithm in handling mixed effects ODEs featuring manifest variables only has been evaluated elsewhere (Donnet & Samson, 2007; Kuhn & Lavielle, 2005), these researchers’ prior work did not explicitly consider the performance of the SAEM in situations where the initial conditions are unknown and differ between subjects. Thus, as distinct from the work of Kuhn and Lavielle (2005) and Donnet and Samson (2007), the proposed modeling framework contributes uniquely to the literature on ODE modeling by (1) including a factor analytic model as a measurement model to enable modeling at the latent variable level, and (2) allowing the means and interindividual differences in the initial conditions of latent variables to be estimated as modeling parameters. Additionally, a simulation study is conducted to evaluate the performance of the estimation procedures under different initial condition specifications, including scenarios where the initial conditions are misspecified and interindividual differences in initial conditions are ignored.

1. Nonlinear Ordinary Differential Equation (ODE) Models with Random Effects and Possibly Unknown Initial Conditions

We consider the problem of fitting linear and nonlinear ODEs with random effects in the structural parameters under situations in which the initial conditions of the ODEs may be unknown. Letting $D\mathbf{x}_i(t)$ denote the differential operator applied on $\mathbf{x}_i(t)$, i.e., $D\mathbf{x}_i(t) = d\mathbf{x}_i(t)/dt$, the nonlinear ODEs of interest take on the general form of

$$\begin{aligned} D\mathbf{x}_i(t) &= \mathbf{f}[\mathbf{x}_i(t), \boldsymbol{\theta}_{f,i}, t], \quad i = 1, \dots, n \\ \boldsymbol{\theta}_{f,i} &= \mathbf{H}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \end{aligned} \quad (1)$$

where i indexes person and t indexes time, $\mathbf{f}(\cdot)$ is a vector of (possibly nonlinear) drift functions; $\mathbf{x}_i(t)$ is an $n_x \times 1$ vector of latent variables of interest at time t ; and $D\mathbf{x}_i(t)$ denotes the corresponding $n_x \times 1$ vector of first derivatives. Note that $\mathbf{x}_i(t)$ may include latent derivative variables needed to define higher-order ODEs. $\boldsymbol{\theta}_{f,i}$ represents a $q \times 1$ vector of person-specific structural parameters of interest that affect the dynamic functions in Eq. (1), expressed as a function of $\boldsymbol{\beta}$, a $p_\beta \times 1$ vector of fixed effects parameters, and \mathbf{b}_i , a $d \times 1$ vector of random effect; \mathbf{H}_i and \mathbf{Z}_i are $q \times p_\beta$ and $q \times d$ design matrices typically seen in the linear mixed effects framework. We further assume that \mathbf{b}_i follows a multivariate normal distribution as $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$, where $\boldsymbol{\theta}_b$ contains all the unknown parameters in $\boldsymbol{\Sigma}_b$.

The initial conditions for the ODEs are denoted as $\mathbf{x}_i(t_{i,1})$, and are specified to be functions of $\boldsymbol{\theta}_{f,i}$. In this way, fixed effects parameters governing the initial conditions are estimated as part of $\boldsymbol{\beta}$, while individual-specific deviations in initial conditions are captured by the random effects in \mathbf{b}_i . Our illustrative application provides a concrete example of one possible way of representing unknown initial conditions across multiple subjects using this formulation.

The latent variables in $\mathbf{x}_i(t_{i,j})$ at discrete time point $t_{i,j}$ are indicated by an $n_y \times 1$ vector of manifest observations assumed to be measured at individual-specific and possibly irregularly spaced time intervals, at $t = t_{i,j}$, $j = 1, \dots, T$, with $\Delta_{i,j} = t_{i,j+1} - t_{i,j}$. The vector of manifest observations is denoted as $\mathbf{y}_i(t_{i,j})$, with

$$\mathbf{y}_i(t_{i,j}) = \boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{x}_i(t_{i,j}) + \boldsymbol{\epsilon}_i(t_{i,j}), \quad (2)$$

where $\boldsymbol{\mu}$ is an $n_y \times 1$ vector of intercepts, $\boldsymbol{\Lambda}$ is an $n_y \times n_x$ factor loading matrix, and $\boldsymbol{\epsilon}_i(t_{i,j})$ denotes multivariate normally distributed measurement error processes such that $E[\boldsymbol{\epsilon}_i(t_{i,j})] = \mathbf{0}$, $E[\boldsymbol{\epsilon}_i(t_{i,j})\boldsymbol{\epsilon}_i(t_{i,j})^T] = \boldsymbol{\Sigma}_\epsilon$, $\boldsymbol{\epsilon}_i(t_{i,j})$ and $\mathbf{x}_i(t_{i,j})$ are independent, and $\boldsymbol{\epsilon}_i(t_{i,j})$ and $\boldsymbol{\epsilon}_i(t_{i,k})$ are independent for $t_{i,j} \neq t_{i,k}$, with a diagonal structure for $\text{Cov}(\boldsymbol{\epsilon}_i(t_{i,j})) = \boldsymbol{\Sigma}_\epsilon$. We also assume generally that when multiple indicators are used to identify a latent factor, at least one factor loading is fixed for identification purposes.

In continuous time, given the initial conditions of the system at any arbitrary time t_0 , one may define a t-advance mapping or evolution function $g(t, \mathbf{x}_i(t_0))$ that moves an initial state to a later state at time t . The evolution function, $g(t, \mathbf{x}_i(t_0))$, can be used to map out the trajectories of all the latent variables in $\mathbf{x}_i(t)$. Once these values are known, the solution to the hypothesized ODE is also known. Thus, one can obtain the solution of the ODE through repeated application of the evolution function, e.g., first from time t_0 to s , and subsequently to t , as $g(t+s, \mathbf{x}_i(t_0)) = g(t, g(s, \mathbf{x}_i(t_0)))$. Thus, if the vector of true latent differences, defined as $\Delta \mathbf{x}_i(t_{i,j}) = \mathbf{x}_i(t_{i,j}) - \mathbf{x}_i(t_{i,j-1})$, are known at a series of discrete time points, $t_{i,j}$, $j = 1, \dots, T$, one can obtain a discrete t-advance mapping as

$$\begin{aligned} \mathbf{x}_i(t_{i,j}) &= \mathbf{x}_i(t_{i,j-1}) + \Delta \mathbf{x}_i(t_{i,j}), \\ &= \mathbf{x}_i(t_{i,1}) + \sum_{k=2}^j \Delta \mathbf{x}_i(t_{i,k}). \end{aligned} \quad (3)$$

With few exceptions, most nonlinear ODEs do not have analytic solutions. One common approach is to use numerical methods such as Euler's or Runge-Kutta methods to obtain approximate t-advance mapping at discrete intervals. In other words, we approximate the vector of true latent differences, $\Delta \mathbf{x}_i(t_{i,j+1}) = \mathbf{x}_i(t_{i,j+1}) - \mathbf{x}_i(t_{i,j})$ using $\mathbf{x}_i^*(t_{i,j})$, namely, numerical interpolations of the latent changes at the next time point based on the hypothesized ODE to yield numerical solution, $\tilde{\mathbf{x}}_i(t_{i,j})$, as

$$\begin{aligned} \tilde{\mathbf{x}}_i(t_{i,j}) &= \tilde{\mathbf{x}}_i(t_{i,j-1}) + \sum_{k=1}^{1/\Delta^*} \mathbf{x}_i^*(t_{i,j-1+k\Delta^*}), \\ &= \mathbf{x}_i(t_{i,1}) + \sum_{k=1}^{(j-1)/\Delta^*} \mathbf{x}_i^*(t_{i,1+k\Delta^*}), \end{aligned} \quad (4)$$

where the numerical latent differences, $\mathbf{x}_i^*(t_{i,j})$, are typically obtained at an equally spaced interval, Δ^* , whose magnitude is considerably smaller than the observed measurement intervals, $\Delta_{i,j}$, to improve the accuracy of the solutions.

We define $t_{i,j*} = t_{i,j-1+k\Delta^*}$ as a time point at which observed measurements might not be available but we are interested in “imputing” the values of the latent changes at this point, $\mathbf{x}_i^*(t_{i,j*})$, to improve estimation accuracy. A variety of numerical solvers can be used to obtain $\mathbf{x}_i^*(t_{i,j*})$. One such example is the second-order Heun’s method, which is implemented as

$$\begin{aligned}\mathbf{x}_i^*(t_{i,j*}) &= \frac{\Delta^*}{2} \{ \mathbf{k}_1(t_{i,j*-\Delta^*}) + \mathbf{k}_2(t_{i,j*}) \}, \\ \mathbf{k}_1(t_{i,j*-\Delta^*}) &= f(\tilde{\mathbf{x}}_i(t_{i,j*-\Delta^*}), \boldsymbol{\theta}_{f,i}, t_{i,j*-\Delta^*}), \\ \mathbf{k}_2(t_{i,j*}) &= f(\tilde{\mathbf{x}}_i(t_{i,j*-\Delta^*}) + \Delta^* \mathbf{k}_1(t_{i,j*-\Delta^*}), \boldsymbol{\theta}_{f,i}, t_{i,j*}).\end{aligned}\quad (5)$$

Conditional on \mathbf{b}_i , if the initial condition variables in $\mathbf{x}_i(t_{i,1})$ are known or can be estimated, then $\tilde{\mathbf{x}}_i(t_{i,j})$ is also known. This reflects the *deterministic* nature of ODEs. The fitted model thus becomes

$$\mathbf{y}_i(t_{i,j}) | \tilde{\mathbf{x}}_i(t_{i,j}) \overset{\text{approx}}{\sim} N(\boldsymbol{\mu} + \boldsymbol{\Lambda} \tilde{\mathbf{x}}_i(t_{i,j}), \boldsymbol{\Sigma}_\epsilon), \quad (6)$$

where the notation $\overset{\text{approx}}{\sim}$ denotes approximately distributed as, used to highlight the fact that the numerical solutions given by $\tilde{\mathbf{x}}_i(t_{i,j})$ contain truncation errors, namely, errors stemming from using the ODE solver to numerically approximate the true solutions of the ODE.¹ It is important to emphasize that the true latent variables, $\mathbf{x}_i(t_{i,j})$, as well as the associated approximation terms, $\mathbf{x}_i^*(t_{i,j})$ and $\tilde{\mathbf{x}}_i(t_{i,j})$, are all functions of $\boldsymbol{\theta}_{f,i}$. Here, we suppress the notational dependency to ease presentation.

Our interest is in estimating the parameters in $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\theta}_\mu^T, \boldsymbol{\theta}_\Lambda^T, \boldsymbol{\theta}_\epsilon^T, \boldsymbol{\theta}_b^T)$ via the SAEM algorithm (Zhu & Gu, 2007). Here, $\boldsymbol{\theta}_\mu$ is a $p_\mu \times 1$ (where $p_\mu \leq n_y$) vector of freed parameters in $\boldsymbol{\mu}$, $\boldsymbol{\theta}_\Lambda$ is a $p_\Lambda \times 1$ vector containing all unknown factor loadings in $\boldsymbol{\Lambda}$, $\boldsymbol{\theta}_\epsilon$ is a $p_\epsilon \times 1$ vector containing all the unknown parameters in $\boldsymbol{\Sigma}_\epsilon$, and $\boldsymbol{\theta}_b$ is a $p_b \times 1$ vector containing all the unknown parameters in $\boldsymbol{\Sigma}_b$.

1.1. Stochastic Approximation Expectation–Maximization (SAEM) Algorithm

Prior to describing the SAEM algorithm, we first introduce some key notations. Let $\mathbf{Y}_i(t_{i,j}) = \{\mathbf{y}_i(t_{i,1}), \dots, \mathbf{y}_i(t_{i,j})\}$, $\mathbf{Y}_i = \{\mathbf{y}_i(t_{i,1}), \dots, \mathbf{y}_i(t_{i,T})\}$, and $\mathbf{Y} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ is the observed data array for all n participants; let $\tilde{\mathbf{X}}_i(t_{i,j}) = \{\tilde{\mathbf{x}}_i(t_{i,1}), \dots, \tilde{\mathbf{x}}_i(t_{i,j})\}$, $\tilde{\mathbf{X}}_i = \{\tilde{\mathbf{x}}_i(t_{i,1}), \dots, \tilde{\mathbf{x}}_i(t_{i,T})\}$, and $\tilde{\mathbf{X}} = \{\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_n\}$ is the array of numerical solutions of the latent variables for all n participants. Further, we denote the augmented complete data array as $\mathbf{Z} = \{\mathbf{Y}, \mathbf{b}\}$. In standard Expectation–Maximization (EM) procedures (Dempster, Laird, & Rubin, 1977), maximum likelihood estimates (MLEs) are obtained by cycling iteratively through an expectation (E)-step and a maximization (M)-step. The E-step typically involves analytically computing terms that appear in a pseudo-loglikelihood function given by the conditional expectation of the complete-data loglikelihood function with respect to the distribution $p(\mathbf{b} | \mathbf{Y}; \boldsymbol{\theta})$. The M-step involves updating the parameter estimates using analytic formulas that serve to maximize the pseudo-loglikelihood function. SAEM differs from conventional E–M algorithms in the use of a stochastic approximation procedure in the E-step, coupled with a gradient-type updating procedure (e.g., the Newton–Raphson and Gauss–Newton algorithms; Ortega, 1990) in the M-step. That is, while the required expectations are analytically intractable, the E-step is made possible by replacing analytic expectations with summary statistics computed using samples drawn from the conditional distribution, $p(\mathbf{b} | \mathbf{Y}; \boldsymbol{\theta})$, by means of Markov chain Monte Carlo (MCMC) procedures. A Newton–Raphson

¹ The local truncation errors of a numerical solver at each time point are equal to $c\Delta_{i,j}^{g+1}$, where g is the order of the ODE solver and c is a vector of constants that depends on elements such as the differentials of the ODEs (for further details see Press, Teukolsky, Vetterling, & Flannery, 2002; Ralston & Rabinowitz, 2001).

algorithm is then used to obtain MLEs of θ in the maximization (M)-step during which a sequence of gain constants, γ , is used to control the degree to which updates of parameter estimates are weighted in subsequent iterations.

In the present context, the complete-data probability density function is given by

$$p(\mathbf{Y}, \mathbf{b}; \theta) = \int p(\mathbf{Y}|\tilde{\mathbf{X}}, \mathbf{b}; \theta) p(\tilde{\mathbf{X}}|\mathbf{b}; \theta) p(\mathbf{b}; \theta) d\tilde{\mathbf{X}}.$$

However, due to the deterministic nature of the ODE, $p(\tilde{\mathbf{X}}|\mathbf{b}; \theta)$ is known conditional on \mathbf{b} .² Thus, the integration over $\tilde{\mathbf{X}}$ vanishes, and the complete-data loglikelihood function reduces to

$$\begin{aligned} \log[p(\mathbf{Y}, \mathbf{b}; \theta)] &\triangleq L(\mathbf{Z}; \theta) = \sum_{i=1}^n \left[L_i(\mathbf{b}; \theta) + \sum_{j=1}^T L_{i,j}(\mathbf{Y}|\mathbf{b}; \theta) \right], \text{ with} \\ L_{i,j}(\mathbf{Y}|\mathbf{b}; \theta) &= -\frac{1}{2} \left\{ (n_y) \log(2\pi) + \log |\boldsymbol{\Sigma}_\epsilon| + [\mathbf{y}_i(t_{i,j}) - \boldsymbol{\mu} - \boldsymbol{\Lambda} \tilde{\mathbf{x}}_i(t_{i,j})]^T \boldsymbol{\Sigma}_\epsilon^{-1} \right. \\ &\quad \left. [\mathbf{y}_i(t_{i,j}) - \boldsymbol{\mu} - \boldsymbol{\Lambda} \tilde{\mathbf{x}}_i(t_{i,j})] \right\}, \text{ and} \\ L_i(\mathbf{b}; \theta) &= -\frac{1}{2} \left\{ d \log(2\pi) + \log |\boldsymbol{\Sigma}_b| + \mathbf{b}_i^T \boldsymbol{\Sigma}_b^{-1} \mathbf{b}_i \right\}, \end{aligned} \quad (7)$$

where the associated score vector and information matrix are denoted as $s_{\mathbf{Z}}(\theta; \mathbf{Z})$ and $\mathbf{I}_{\mathbf{Z}}(\theta; \mathbf{Z})$. Detailed analytical forms of $s_{\mathbf{Z}}(\theta; \mathbf{Z})$ and $\mathbf{I}_{\mathbf{Z}}(\theta; \mathbf{Z})$ are included in Appendix 1.

To use any gradient-type algorithm such as the Newton–Raphson (Ortega, 1990) requires the score vector and the information (or negative Hessian) matrix of the loglikelihood function, denoted as $s_{\mathbf{Y}}(\theta; \mathbf{Y})$ and $\mathbf{I}_{\mathbf{Y}}(\theta; \mathbf{Y})$, respectively. They can be obtained as (Dempster et al., 1977; Louis, 1982)

$$\begin{aligned} s_{\mathbf{Y}}(\theta; \mathbf{Y}) &= \frac{\partial L(\mathbf{Y}; \theta)}{\partial \theta} = E(s_{\mathbf{Z}}(\theta; \mathbf{Z})|\mathbf{Y}; \theta), \text{ and} \\ \mathbf{I}_{\mathbf{Y}}(\theta; \mathbf{Y}) &= -\frac{\partial^2 L(\mathbf{Y}; \theta)}{\partial \theta \partial \theta^T} \\ &= E \left[\mathbf{I}_{\mathbf{Z}}(\theta; \mathbf{Z}) - s_{\mathbf{Z}}(\theta; \mathbf{Z}) s_{\mathbf{Z}}(\theta; \mathbf{Z})^T \middle| \mathbf{Y}; \theta \right] + s_{\mathbf{Y}}(\theta; \mathbf{Y}) s_{\mathbf{Y}}(\theta; \mathbf{Y})^T, \end{aligned} \quad (8)$$

where $E(\cdot|\mathbf{Y}, \theta)$ denotes expectation taken with respect to the conditional distribution of $p(\mathbf{b}|\mathbf{Y}; \theta)$. Once the score function and information matrix, $s_{\mathbf{Y}}(\theta; \mathbf{Y})$ and $\mathbf{I}_{\mathbf{Y}}(\theta; \mathbf{Y})$, are available, one can then obtain updated estimates of the parameter vector, θ , at iteration m using any gradient-type algorithm, such as the Newton–Raphson as (Gu & Zhu, 2001; Zhu & Gu, 2007; Ortega, 1990)

$$\theta^{(m)} = \theta^{(m-1)} + [\mathbf{I}_{\mathbf{Y}}(\theta^{(m-1)}; \mathbf{Y})]^{-1} s_{\mathbf{Y}}(\theta^{(m-1)}; \mathbf{Y}). \quad (9)$$

While the classical EM algorithm and the Newton–Raphson procedure shown in (9) both serve to provide iterative updates of the parameter estimates to yield ML estimates for which $s_{\mathbf{Y}}(\theta; \mathbf{Y}) = \mathbf{0}$, the Newton–Raphson algorithm and other related gradient-type algorithms typically outperform

² In contrast, in cases involving SDEs, $p(\tilde{\mathbf{X}}|\mathbf{b}; \theta)$ is not fixed even when \mathbf{b} is known and there is considerable increase in estimation complexity.

the classical EM in terms of convergence speed (Dembo & Zeitouni, 1986; Ortega, 1990; Singer, 1995).

Based on Eq. (7) and its constituent elements shown in Appendix 1, taking the expectations of $s_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})$ and $\mathbf{I}_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})$ with respect to the distribution $p(\mathbf{b}|\mathbf{Y}; \boldsymbol{\theta})$ requires the computation of terms such as $E[\tilde{\mathbf{x}}(t_{i,j})|\mathbf{Y}; \boldsymbol{\theta}^{(m-1)}]$, $E\left(\frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})}{\partial \theta_{f,i}}|\mathbf{Y}; \boldsymbol{\theta}^{(m-1)}\right)$, $E[\tilde{\mathbf{x}}(t_{i,j})\tilde{\mathbf{x}}_i(t_{i,j})^T|\mathbf{Y}; \boldsymbol{\theta}^{(m-1)}]$, and $E\left(\frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T}{\partial \theta_{f,i}}\mathbf{W}^*\tilde{\mathbf{x}}(t_{i,j})|\mathbf{Y}; \boldsymbol{\theta}^{(m-1)}\right)$, where \mathbf{W}^* is a function involving vectors/matrices of modeling parameters such as $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}_\epsilon$. These terms, in turn, require integration over $p(\mathbf{b}|\mathbf{Y}; \boldsymbol{\theta})$. Because \mathbf{b}_i is involved in the nonlinear $f(\cdot)$ in the computation of $\tilde{\mathbf{x}}(t_{i,j})$, such integration is analytically formidable. In the present article, we chose to perform a variation of the EM algorithm—the SAEM algorithm—for parameter estimation purposes.

E-step with stochastic approximation In the E-step, a MCMC technique of choice is used to simulate a sequence of random draws from the conditional distribution, $p(\mathbf{b}|\mathbf{Y}; \boldsymbol{\theta})$. For reasons detailed in Appendix 2, $p(\mathbf{b}|\mathbf{Y}; \boldsymbol{\theta})$ is non-standard, and sampling from this distribution cannot be performed directly by means of Gibbs sampling. We thus adopt the Metropolis–Hasting sampling procedure described in Appendix 2 to simulate samples of \mathbf{b}_i probabilistically from an alternative proposal distribution that does have a standard form to obtain $\mathbf{Z}_k^{(m)} = (\mathbf{Y}, \mathbf{b}_k^{(m)})$, which is then used in the E-step to compute the summary statistics

$$\begin{aligned}\bar{s}_{\mathbf{Z}}^{(m)} &= \frac{1}{N_m} \sum_{k=1}^{N_m} s_{\mathbf{Z}}(\boldsymbol{\theta}^{(m-1)}; \mathbf{Z}_k^{(m)}), \quad \bar{\mathbf{S}}_{\mathbf{Z}}^{(m)} = \frac{1}{N_m} \sum_{k=1}^{N_m} s_{\mathbf{Z}}(\boldsymbol{\theta}^{(m-1)}; \mathbf{Z}_k^{(m)}) s_{\mathbf{Z}}(\boldsymbol{\theta}^{(m-1)}; \mathbf{Z}_k^{(m)})^T \text{ and} \\ \bar{\mathbf{I}}_{\mathbf{Z}}^{(m)} &= \frac{1}{N_m} \sum_{k=1}^{N_m} \mathbf{I}_{\mathbf{Z}}(\boldsymbol{\theta}^{(m-1)}; \mathbf{Z}_k^{(m)}),\end{aligned}\quad (10)$$

where $s_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})$ and $\mathbf{I}_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})$ are defined, respectively, in Eqs. (17) and (18) of Appendix 1. These summary statistics are then used in the M-step to update the parameter estimates in $\boldsymbol{\theta}$.

M-step In the M-step, the goal is to obtain updated parameter estimates, $\boldsymbol{\theta}^{(m)}$, for the m th iteration using a modified Newton–Raphson procedure involving $\bar{s}_{\mathbf{Z}}^{(m)}$ and current estimates of the information matrix, $\mathbf{I}_{\mathbf{Y}}^{(m)}(\boldsymbol{\theta}; \mathbf{Y})$, abbreviated below as $\mathbf{I}_{\mathbf{Y}}^{(m)}$. Doing so in turn requires (see Eq. 9) current estimates of the score function, $s_{\mathbf{Y}}(\boldsymbol{\theta}; \mathbf{Y})^{(m)}$, abbreviated below as $s_{\mathbf{Y}}^{(m)}$, current estimates of $E[\mathbf{I}_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})|\mathbf{Y}, \boldsymbol{\theta}]$, denoted as $E_{\mathbf{I}}^{(m)}$, and estimates of $E[s_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})s_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})^T|\mathbf{Y}, \boldsymbol{\theta}]$, denoted as $E_{\mathbf{S}}^{(m)}$. These elements are computed as

$$\begin{aligned}s_{\mathbf{Y}}^{(m)} &= s_{\mathbf{Y}}^{(m-1)} + \gamma^{(m)}[\bar{s}_{\mathbf{Z}}^{(m)} - s_{\mathbf{Y}}^{(m-1)}], \\ \mathbf{I}_{\mathbf{Y}}^{(m)} &= E_{\mathbf{I}}^{(m)} - E_{\mathbf{S}}^{(m)} + (s_{\mathbf{Y}}^{(m)})(s_{\mathbf{Y}}^{(m)})^T, \\ E_{\mathbf{S}}^{(m)} &= E_{\mathbf{S}}^{(m-1)} + \gamma^{(m)}[\bar{\mathbf{S}}_{\mathbf{Z}}^{(m)} - E_{\mathbf{S}}^{(m-1)}], \\ E_{\mathbf{I}}^{(m)} &= E_{\mathbf{I}}^{(m-1)} + \gamma^{(m)}[\bar{\mathbf{I}}_{\mathbf{Z}}^{(m)} - E_{\mathbf{I}}^{(m-1)}], \\ \boldsymbol{\theta}^{(m)} &= \boldsymbol{\theta}^{(m-1)} + \gamma^{(m)}[\mathbf{I}_{\mathbf{Y}}^{(m)}]^{-1}s_{\mathbf{Y}}^{(m)},\end{aligned}\quad (11)$$

where $\gamma^{(m)}$ is a gain constant that controls the degree to which new estimates are weighted at iteration m in comparison with the estimates from iteration $m - 1$. The gain constant is a control parameter that is modified in two stages to (1) prevent the estimation algorithm from settling too quickly into local minima in earlier iterations (stage 1; for iteration $m = 1, \dots, K_1$); and (2) help speed convergence toward a final set of parameter estimates during the later iterations (stage 2;

for iteration $m = K_1 + 1, \dots, K_2$). The use of a gain function is the key feature that distinguishes the SAEM from the stochastic and Monte Carlo EM procedures (Diebolt & Celeux, 1993; Lee & Song, 2003).³

The SAEM algorithm alternates between the E-step and the M-step until some predefined convergence criteria have been met (Zhu & Gu, 2007).⁴ An offline averaging procedure is implemented concurrently as the scoring algorithm in stage 2 of the SAEM. At the conclusion of stage 1 (i.e., at the K_1 th iteration), the averaging procedure is initiated with $\tilde{s}_{\mathbf{Y}}^{(1)} = s_{\mathbf{Y}}^{(K_1)}$, $\tilde{E}_{\mathbf{S}}^{(1)} = E_{\mathbf{S}}^{(K_1)}$, $\tilde{E}_{\mathbf{I}}^{(1)} = E_{\mathbf{I}}^{(K_1)}$, $\tilde{\mathbf{I}}_{\mathbf{Y}}^{(1)} = \mathbf{I}_{\mathbf{Y}}^{(K_1)}$, and $\tilde{\boldsymbol{\theta}}^{(1)} = \boldsymbol{\theta}^{(K_1)}$. These offline estimates are subsequently updated as

$$\begin{aligned}\tilde{s}_{\mathbf{Y}}^{(m)} &= \tilde{s}_{\mathbf{Y}}^{(m-1)} + (s_{\mathbf{Y}}^{(m)} - \tilde{s}_{\mathbf{Y}}^{(m-1)})/m, & \tilde{E}_{\mathbf{S}}^{(m)} &= \tilde{E}_{\mathbf{S}}^{(m-1)} + (E_{\mathbf{S}}^{(m)} - \tilde{E}_{\mathbf{S}}^{(m-1)})/m, \\ \tilde{E}_{\mathbf{I}}^{(m)} &= \tilde{E}_{\mathbf{I}}^{(m-1)} + (E_{\mathbf{I}}^{(m)} - \tilde{E}_{\mathbf{I}}^{(m-1)})/m, & \tilde{\mathbf{I}}_{\mathbf{Y}}^{(m)} &= \tilde{\mathbf{I}}_{\mathbf{Y}}^{(m-1)} + (\mathbf{I}_{\mathbf{Y}}^{(m)} - \tilde{\mathbf{I}}_{\mathbf{Y}}^{(m-1)})/m, \\ \tilde{\boldsymbol{\theta}}^{(m)} &= \tilde{\boldsymbol{\theta}}^{(m-1)} + (\boldsymbol{\theta}^{(m)} - \tilde{\boldsymbol{\theta}}^{(m-1)})/m.\end{aligned}\quad (13)$$

Theoretically, this averaging procedure automatically leads to an optimal convergence without estimating the information matrix (Polyak, 1990; Polyak and Juditski, 1992). Under some conditions, the offline average $(\tilde{\boldsymbol{\theta}}^{(m)}, \tilde{s}_{\mathbf{Y}}^{(m)})$ converges to $(\hat{\boldsymbol{\theta}}_{MLE}, \hat{s}_{MLE}(\boldsymbol{\theta}; \mathbf{Y}))$ almost surely, as $K_1 + K_2$ approaches infinity (Zhu & Gu, 2007). Finally, at convergence, we use the offline average, $(\tilde{\boldsymbol{\theta}}^{(K_2)}, \tilde{\mathbf{I}}_{\mathbf{Y}}^{(K_2)})$, as our final estimate of $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{I}}_{\mathbf{Y}}(\boldsymbol{\theta}; \mathbf{Y}))$, where square roots of the diagonal elements of $\hat{\mathbf{I}}_{\mathbf{Y}}(\boldsymbol{\theta}; \mathbf{Y})$ are used as the standard error (SE) estimates of the parameters.

2. Illustrative Application

To illustrate the empirical utility of the proposed approach, we reanalyzed a set of previously published data (Carels, Blumenthal, & Sherwood, 2000; Sherwood, Steffen, Blumenthal, Kuhn, & Hinderliter, 2002) involving 172 employed men and women, aged 25–45 years, who participated in the Duke Biobehavioral Investigation of Hypertension study. The sample comprised 96 participants with normal blood pressure (BP), 41 with high normal BP, and 35 with stage 1 hypertension. Clinic systolic blood pressure that was greater than 180 mmHg or diastolic blood pressure that was greater than 100 mmHg, use of cardiovascular medications, and use of tobacco products were specified as exclusion criteria.

The AccuTracker II ABP Monitor (Suntech AccuTracker II, Raleigh, NC) was worn for approximately 24 h to measure BP and heart rate noninvasively. It was programmed to take four measurements hourly at random intervals ranging from 12 to 28 minutes apart. Participants were instructed to follow their normal schedule and to complete a diary entry indicating posture, activity, location, positive affect, and negative affect at each reading. The same procedure was

³ In the present context, we specify the gain constant to be $\gamma^{(m)} = a_2/(m^{a_1} + a_2 - 1)$, $m = 1, \dots, K_1 + K_2$, where the real number a_1 and the integer a_2 are preassigned. In stage 1, a_1 and a_2 are selected such that the gain constant assumes some large values to prevent the SAEM algorithm from settling into local minima too quickly. In stage 2, the gain constant is slowly tapered toward zero to allow the algorithm to stabilize toward a final set of estimates (e.g., by setting $a_1 \in (.5, 1]$ to be close to 1, and a_2 to be a small integer, say, $a_2 = 2$). The transition from stage 1 to 2 is governed by another predefined criterion function (for details see Gu & Zhu, 2001; Zhu & Gu, 2007).

⁴ In the present study, we define the stopping rule to be

$$K_2 = \inf \left\{ m : \tilde{s}_{\mathbf{Y}}^{(m)} [\tilde{\mathbf{I}}_{\mathbf{Y}}^{(m)}]^{-1} \tilde{s}_{\mathbf{Y}}^{(m)} + \text{tr} \left\{ [\tilde{\mathbf{I}}_{\mathbf{Y}}^{(m)}]^{-1} \hat{\boldsymbol{\Sigma}} \right\} / m \leq \text{some small constant} \right\}. \quad (12)$$

$\hat{\boldsymbol{\Sigma}}$ denotes an estimate of the covariance matrix of Monte Carlo error. In practice, we used the sample covariance matrix of $\bar{s}_{\mathbf{Z}}^{(m)}$ as a rough estimate of $\hat{\boldsymbol{\Sigma}}$.

implemented in the evening waking hours. Sleep was defined by diary activity ratings, which included an indication of “going to sleep.” The monitor was programmed to take only two readings hourly during sleeping hours, customized to the participants’ sleep habits.

The dynamics of individuals’ ambulatory BP and other related cardiovascular measures during a typical workday were the focus of our modeling example. In the study of 24-h ambulatory BP, BP dipping, an important prognostic indicator of cardiovascular morbidity and mortality, is derived by subtracting the mean nighttime sleep BP from the mean daytime waking BP (Sherwood et al., 2001, 2002). Consequently, the intensive BP readings collected throughout the study period (which typically exceed 50 measurements per participant) are aggregated and reduced to only mean-level information (Pickering, Shimbo, & Haas, 2006). Even in the few studies where within-subject linkages between ambulatory BP and related risk factors were explored without further data aggregation (e.g., Carels, 2000), emphasis was placed almost exclusively on examining *concurrent* linkages between multiple processes as opposed to lagged interdependencies among change processes. In this way, within-subject correlations between successive repeated measurements may be allowed (e.g., via Generalized Estimating Equations as in Carels, 2000) but the precise mechanisms of change (e.g., interdependencies among change processes) as well as related interindividual differences therein are not modeled explicitly. In the present analysis, we used the modified Van der Pol oscillator model to evaluate individual differences in the dynamics of diurnal cardiovascular fluctuations. In particular, we examined whether negative and positive emotion ratings throughout the day would help predict individual differences in the amplification of diurnal cardiovascular fluctuations.

Brown and colleagues (Brown & Luithardt, 1999; Brown et al., 2000) adapted the well-known Van der Pol oscillator equations to model human circadian data. In the present context, constrained by the lack of sufficient time points to capture slower circadian cycles and deviations therein, we considered a modified Van der Pol oscillator model as

$$\begin{aligned}
 D \begin{bmatrix} x_{i1}(t) \\ x_{i2}(t) \end{bmatrix} &= \begin{bmatrix} x_{i2}(t) \\ -\left(\frac{2\pi}{\gamma}\right)^2 x_{i1}(t) + \zeta_i [1 - x_{i1}^2(t)] x_{i2}(t) \end{bmatrix}, \quad \mathbf{x}_i(t_{i,1}) = \begin{bmatrix} x_{i1}(t_{i,1}) \\ x_{i2}(t_{i,1}) \end{bmatrix}, \\
 \boldsymbol{\theta}_{f,i} &= \begin{bmatrix} \zeta_i \\ x_{i1}(t_{i,1}) \\ x_{i2}(t_{i,1}) \end{bmatrix} = \begin{bmatrix} 1 & u_{1i} & u_{2i} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \zeta_0 \\ \zeta_1 \\ \zeta_2 \\ \mu_{x1} \\ \mu_{x2} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{\zeta,i} \\ b_{x1,i} \\ b_{x2,i} \end{bmatrix} \\
 \begin{bmatrix} y_{i1}(t_{i,j}) \\ y_{i2}(t_{i,j}) \\ y_{i3}(t_{i,j}) \end{bmatrix} &= \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \end{bmatrix} \begin{bmatrix} x_{i1}(t_{i,j}) \\ x_{i2}(t_{i,j}) \end{bmatrix} + \begin{bmatrix} \epsilon_{i1}(t_{i,j}) \\ \epsilon_{i2}(t_{i,j}) \\ \epsilon_{i3}(t_{i,j}) \end{bmatrix}, \tag{14}
 \end{aligned}$$

where $x_{i1}(t)$ represents the level of a dependent variable of interest (in our case, cardiovascular reactivity) and $x_{i2}(t)$ is its corresponding first derivative. In the present application, $x_{i1}(t)$ was indicted by three manifest variables, namely, systolic blood pressure, diastolic blood pressure, and heart rate, with μ_1 – μ_3 as their respective intercepts. ζ_i is a general damping or amplification parameter that governs the oscillation amplitude of person i , and $(2\pi/\gamma)^2$ is the squared frequency of cardiovascular reactivity in radians in the absence of damping or amplification, with γ representing the period of human circadian rhythm, fixed at 2 (i.e., corresponding to a period of 24 h) in the current context.⁵ The amount of damping (or amplification) is further moderated

⁵ This decision was made because there were insufficient repeated measurements to estimate this parameter accurately, especially for individuals with a diurnal cycle that is longer than 24 h, or individuals with less than 24 h worth of measurements.

by a quadratic term in $x_{i1}(t)$. Formulated this way, the amplification (if $\zeta_i > 0$) at small values of $x_{i1}(t)$ is expected to turn into massive damping at extreme values of $x_{i1}(t)$ due to the quadratic term, $x_{i1}^2(t)$. This yields a system that is slow to rise to its peak, followed by damping (manifested as a pronounced drop in amplitude) when it hits extreme values. In contrast, if ζ_i is negative, this yields a system with damping dynamics at small values of $x_{i1}(t)$, and pronounced amplification at extreme values of $x_{i1}(t)$. If the initial level and rate of change of the system are extreme (i.e., far away from zero), the quadratic term, $x_{i1}^2(t)$, would dominate the system's dynamics by moderating and magnifying the system's amplification, resulting in an explosive system—a scenario that is rare in CV dynamics. Since we expected ζ_i to be positive, we refer to this parameter as the amplification parameter throughout. In this illustrative example, ζ_i is further expressed as a function of a group intercept, ζ_0 , two person-specific covariates, u_{1i} and u_{2i} (with fixed effects parameters, ζ_1 and ζ_2 , respectively), and $b_{\zeta,i}$, person i 's deviation in ζ_i that is not accounted for by other fixed effects terms. As an example of some of the between-person heterogeneities in dynamics one may obtain from the hypothesized model, simulated trajectories of $x_{i1}(t_{i,j})$ using different values of ζ and initial conditions are shown in Figure 1a. The two trait-level covariates used in the present context to predict individual differences in the amplification parameter were (overtime) aggregate ratings of negative emotion (NE), obtained by averaging the participants' responses over three items: *stress*, *anger*, and *tense*; and positive emotion (PE), obtained by averaging participants' responses over the items *happy* and *in control*.

The present application also serves to demonstrate how the proposed modeling framework can be used to represent the unknown initial conditions at time 1, $x_{i1}(t_{i,1})$ and $x_{i2}(t_{i,1})$, as part of $\theta_{f,i}$. The formulation adopted in (14) dictates that individual i 's initial conditions at time 1, namely, $x_{i1}(t_{i,1})$ and $x_{i2}(t_{i,1})$, follow a multivariate normal distribution with mean vector $[\mu_{x1}\mu_{x2}]^T$, and a covariance matrix composed of the lower 2×2 submatrix of Σ_b . We assume that the random effects covariance matrix conforms to the structure

$$\Sigma_b = \begin{bmatrix} \sigma_{b_\zeta}^2 & & \\ 0 & \sigma_{b_{x1}}^2 & \\ 0 & \sigma_{b_{x1,x2}} & \sigma_{b_{x2}}^2 \end{bmatrix}, \quad (15)$$

where the person-specific deviation in ζ_i , $b_{\zeta,i}$, is assumed to be uncorrelated with other person-specific deviations in initial conditions, while $b_{x1,i}$ and $b_{x2,i}$ are allowed to covary because interindividual differences in initial level and first derivative are often expected to show non-negligible associations. This particular way of structuring the initial conditions of a dynamic model as unknown parameters to be estimated has rarely been explicitly utilized in ODE modeling, but is commonly adopted in discrete-time state-space and time series models (e.g., Harvey & Souza, 1987; Chow, Ho, Hamaker, & Dolan, 2010; Du Toit & Browne, 2001), as well as growth curve-type models (Meredith & Tisak, 1990; McArdle & Hamagami, 2001). In addition, the presence of statistically significant interindividual differences in the amplification parameter and the two initial latent variables was deduced in the present context from the statistical significance of the variance parameters in Σ_b , namely, by means of a Wald test. A possible alternative to the Wald test will be highlighted in the Discussion section.

After removing data from individuals who contributed less than 50 readings, 168 participants were retained for model fitting purposes. Among these participants, the available measurement occasions ranged from 58 to 98 time points within each individual. Time was rescaled such that one unit of time corresponded to 12 h, with $\Delta_{i,j}$ ranging from 0.001 to 0.48. We fitted the multiple-indicator Van der Pol oscillator model to the ambulatory cardiovascular data. Several preliminary data treatment steps were performed prior to model fitting. First, we observed that substantial interindividual differences were present in the means of these indicator variables

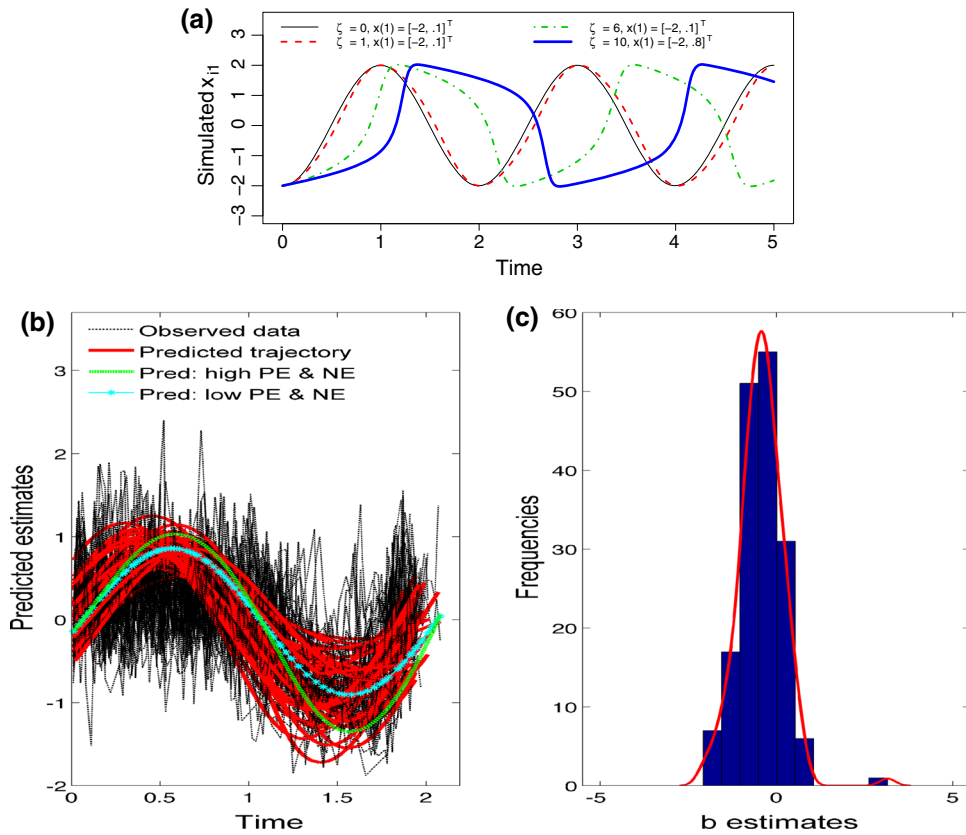


FIGURE 1.

a Simulated trajectories of $x_{i1}(t_{i,j})$ using different values of ζ and initial conditions. The values shown in $x(1)$ are $x_{i1}(t_{i,1})$ and $x_{i2}(t_{i,1})$, respectively. **b** Observed composite cardiovascular data and estimates of the latent cardiovascular reactivity obtained using the Van der Pol oscillator model. Time corresponds to time since each participant's first measurement, as opposed to clock time. **c** Histogram of the random effect estimates of all participants. *Observed data* = composite cardiovascular measure of each participant on each occasion; *Predicted trajectory* = $\tilde{x}_i(t_{i,j})$ generated using the final $\hat{\theta}$ from the SAEM; *Pred: high PE & NE* = $\tilde{x}_i(t_{i,j})$ for a hypothetical individual with PE and NE that were 2 standard deviations higher than the sample average and no person-specific deviations in the amplification parameter and initial conditions; *Pred: low PE & NE* = $\tilde{x}_i(t_{i,j})$ for a hypothetical individual with PE and NE that were 2 standard deviations lower than the sample average and no person-specific deviations in the amplification parameter and initial conditions.

(which had to be modeled by allowing for random effects in the intercepts $\mu_1 - \mu_3$). Because such interindividual differences were not the focus of our illustrative model, we removed these differences by subtracting each individual's mean on each variable from the corresponding time series and used the residual scores for subsequent model fitting. Thus, the intercept parameters, while freed to be estimated as parameters in the present context, were expected to take on values that were close to zero. Second, to remove arbitrary scale differences across the three indicator variables while preserving potential interindividual differences in the amplification parameter, ζ_i , we standardized each individual's time series using the group standard deviation of each indicator variable.

Third, there were substantial individual differences in when the participants took their first measurements of the day, with the first time point ranging in clock time from 7:07 am to 4.82 pm. To eliminate confounds due to individual differences in lifestyle, the participants' dynamics were modeled in terms of time since each participant's first measurement, as opposed to clock time.

Finally, the sparse measurements at night, coupled with the dense daytime measurements, gave rise to highly irregularly spaced time intervals (with $\Delta_{i,j}$ ranging from 0.001 to 0.48). If these time intervals were used as they were, some of the larger time intervals would lead to very large approximation errors regardless of the choice of the ODE solvers, thereby jeopardizing the solvers' numerical stability; the smallest time intervals were so much smaller by comparison that using them directly in the ODE solver would greatly increase computational costs. We adopted some strategies to strike a balance between numerical stability/modeling accuracy and computational time. That is, to improve the numerical stability of the ODE solver, missing data were inserted at the interval of $\Delta_{i,j} = 0.01$ to avoid the need to interpolate over large time intervals. Additionally, we aggregated data that were too densely measured to yield a minimum $\Delta_{i,j}$ of 0.01. This essentially yielded a set of equally spaced data for model fitting purposes. We note that the proposed estimation framework can, in principle, handle irregularly spaced and person-specific $\Delta_{i,j}$. While the first step is needed to improve the numerical stability of the estimation procedures, the latter step is not needed and was implemented simply to reduce computational costs. In practice, the interpolation intervals used in deriving numerical ODE solutions are almost always smaller than the crudest empirically observed measurement intervals. If computational time is not a constraint, then a smaller interpolation interval can help improve numerical accuracy under most circumstances. However, depending on the dynamics of the system, it may not be computationally efficient to always set the interpolation interval to the smallest possible time step. Here, our simulation study showed that interpolating with $\Delta_{i,j} = 0.01$ as the smallest time step in the presence of irregularly spaced time intervals yielded reasonable estimates for the model considered in the present article.

A plot of the composite scores obtained by averaging the participants' rescaled residual scores on the three indicator variables is shown in Figure 1b. Relatively clear diurnal trends can be seen from the observed scores, with the participants' cardiovascular data rising to their peaks and staying at relatively high levels throughout the 12 h (i.e., from time = 0 to about 1) after their first measurements. Declines in cardiovascular levels became evident after $t = 1.2$, with the lowest "dipping" occurring at approximately $t = 1.5$ (corresponding to approximately 18 h after the first measurements).

Parameter estimates obtained from model fitting are shown in Table 1, and the predicted trajectories, $\tilde{x}_i(t_{i,j})$, generated using the final estimates $\hat{\theta}$ from the SAEM, are plotted in Figure 1b. All parameters, except for the interindividual variance in initial condition for the first derivative, σ_{b,x_2}^2 , were significantly different from zero. This indicates that substantive interindividual differences were only present in the initial cardiovascular level at the first time point, but not in its first derivative. The fixed effects associated with PE and NE were both positive and statistically different from 0. The sign of these fixed effects indicated that overall, heightened PE, as well as NE, was found to increase the amplification magnitude of an individual's diurnal cardiovascular oscillations. Such oscillations, characterized by a slow buildup followed by a sudden discharge to relax the stress accumulated during the buildup (Strogatz, 1994, p. 212), appeared to be amplified by emotions of high intensity, regardless of valence. Thus, an individual who experienced heightened PE as well as NE over the course of the study period would show greater build up as well as discharge compared to individuals who did not show high PE and/or NE during the same period. For illustrative purposes, we plotted the trajectories of a hypothetical individual with high PE and NE (defined as 2 standard deviations above the sample average) and low PE and NE (defined as 2 standard deviations below the sample average) in Figure 1b. The fixed effect coefficient for NE was slightly smaller in standardized value compared to that for PE, which is in contrast to commonly held beliefs regarding the close linkages between NE and cardiovascular health. This may be related, however, to differences in the roles of the two emotions in affecting the buildup vs. the discharge phases of cardiovascular dynamics. For instance, while heightened NE may lead to greater buildup of cardiovascular

TABLE 1.
Parameter estimates from fitting the modified Van der Pol Oscillator Model to empirical data.

Parameters	Estimates (SE)
ζ_0	0.49 (0.01)
ζ_1 for NE	0.08 (0.03)
ζ_2 for PE	0.10 (0.02)
μ_{x1}	-0.17 (0.004)
μ_{x2}	2.55 (0.002)
μ_1	-0.11 (0.01)
μ_2	-0.04 (0.01)
μ_3	-0.04 (0.01)
λ_{21}	0.37 (0.01)
λ_{31}	0.44 (0.01)
$\sigma_{\zeta_1}^2$	0.63 (0.01)
$\sigma_{\zeta_2}^2$	0.45 (0.01)
$\sigma_{\zeta_3}^2$	0.476 (0.01)
$\sigma_{b_1}^2$	0.24 (0.02)
$\sigma_{b_{x1}}^2$	0.28 (0.04)
$\sigma_{b_{x2}}^2$	0.19 (0.10)
$\sigma_{b_{x1,x2}}$	0.19 (0.03)

activities in the daytime, it may also be associated with attenuated BP nighttime dipping. Such reversal in damping effects is not captured by the fitted version of the Van der Pol oscillator. A possible modification is to incorporate regime-switching (Chow, Grimm, Guillaume, Dolan, & McArdle, 2013; Chow & Zhang, 2013) or multiphase (Cudeck & Klebe, 2002) extensions wherein the amplification parameter is allowed to show phase-dependent relationships on the covariates.

Several other observations can be noted from the modeling results. First, substantial interindividual differences in cardiovascular amplification remained after the effects of the two covariates had been accounted for. Such differences gave rise to further deviations in the amplitude of diurnal fluctuations (see trajectories of the participants' $\tilde{x}_i(t_{i,j})$ in Figure 1b, generated using the participants' own random effect estimates as shown in Figure 1c). Second, despite the use of residual scores for model fitting purposes, the three intercept parameters were estimated to be negative and significantly different from zero. Third, the factor loadings for the second and third indicators, namely, diastolic BP and heart rate, were low compared to the loading for systolic BP, which was fixed at 1.0 for identification purposes. These latter findings suggest some possible inadequacies of the multivariate Van der Pol oscillator model in capturing the dynamics of the observed data, particularly differences in the diurnal dynamics of the three indicators.

Given the short time series lengths, the relatively small sample size, and other data-related constraints, the results reported here have to be interpreted with caution. For instance, the data collected barely spanned one complete cycle. Thus, there were insufficient time points (and importantly, number of replications in cycle) to clearly distinguish the nature of the oscillations and amplification/damping evidenced by the system over time. In particular, relatively few time points were available during sleeping hours—the time during which the heavy damping as predicted by the Van der Pol oscillator model was supposed to unfold. Nevertheless, our illustration demonstrated the potential promises of the proposed modeling approach and the ways in which future studies of diurnal cardiovascular patterns may be adapted to effectively utilize this approach.

3. Simulation Study

The performance of the SAEM algorithm in handling dynamic models featuring manifest variables only has been evaluated elsewhere (Donnet & Samson, 2007). The novel contributions of this article lie in presenting an alternative formulation that explicitly captures the dynamics of a system and interindividual differences in initial conditions at the latent level. We conducted a simulation study to specifically assess the performance of the SAEM with respect to these novel features. Unlike previous studies that assessed the performance of ODE modeling techniques using small and equally spaced time intervals (e.g., with $\Delta_{i,j} = 0.001$ for all i and j), we used person-specific, irregularly spaced time intervals that mirror those observed in empirical studies.

We used a fourth-order Runge Kutta solver to generate the true latent differences at the population level, while the second-order Heun's method summarized in Eq. (5) was used for model fitting purposes. Three manifest indicators were used to identify $x_{i1}(t)$, with $\Sigma_{\epsilon} = 0.5\mathbf{I}_3$. The first factor loading was set to unity for identification purposes. We set the period of oscillation, γ , to 0.8 (fixed and not estimated), $\zeta_0 = 3$, $\zeta_1 = 0.5$, $\zeta_2 = 0.5$, $\lambda_{21} = 0.7$, $\lambda_{31} = 1.2$, and $\mu_1 = \mu_2 = \mu_3 = 0$. The covariates u_{1i} and u_{2i} were both simulated from a uniform distribution over the interval $[0, 5]$.

Two sample size configurations were considered: (1) $n = 200$ and $T = 300$, with $\Delta_{i,j}$ ranging between 0.005 and 0.07, and (2) $n = 200$ and $T = 150$, with $\Delta_{i,j}$ ranging between 0.006 and 0.161. We also considered two variations of true initial conditions. In the first condition, the initial conditions at time $t_{i,1}$ were set to the constant values, $\mu_{x1} = \mu_{x2} = 1.0$ for all individuals, with the true Σ_b specified to be

$$\Sigma_b = \begin{bmatrix} 0.5 & & \\ 0 & 0 & \\ 0 & 0 & 0 \end{bmatrix}.$$

This condition is denoted as the condition with fixed initial conditions.

For the second condition (denoted as the condition with random initial conditions), we considered initial conditions that conformed to a multivariate normal distribution as follows. We assumed that the process conformed to the hypothesized ODE model both prior to, and after, the first available measurements. To implement this scenario, we began simulating data using the hypothesized ODE fifty time points prior to the first retained measurement occasion, starting at $t_{i,1}$ with $\mu_{x1} = \mu_{x2} = 1.0$ and

$$\Sigma_b = \begin{bmatrix} 0.5 & & \\ 0 & 1.0 & \\ 0 & 0.3 & 1.0 \end{bmatrix}. \quad (16)$$

Then, at $j = 51$, we centered each individual's latent variable scores at the 51st time point using the across-individual means (thus resetting μ_{x1} and μ_{x2} to zero) and retained data from the 51st time point and beyond for model fitting purposes. Given the deterministic nature of the Van der Pol model, the variances and covariances among the true \mathbf{b}_i estimates in the retained data should still mirror the values shown in (16) at the population level. An alternative way of generating data with multivariate normally distributed initial conditions is to simply draw values of $x_{i1}(t_{i,1})$ and $x_{i2}(t_{i,1})$ from a multivariate normal distribution with mean vector $[\mu_{x1} \mu_{x2}]'$ and an arbitrary choice of Σ_b . However, doing so dictates that the dynamics of the ODE system at time $t_{i,1}$ need not follow the dynamics of the system at other remaining time points. We adopted our specification to enforce the assumption of (overtime) homogeneity in dynamics.

Data generated using the two true initial conditions (fixed vs. random) were matched with two fitted initial condition specifications, namely, one where μ_{x1} and μ_{x2} were fixed at the constant value of 1.0 (i.e., with fixed initial conditions), and another one where μ_{x1} , μ_{x2} , $\sigma_{b_{x1}}^2$, $\sigma_{b_{x1,x2}}$, and $\sigma_{b_{x2}}^2$ were all estimated as modeling parameters (i.e., with random initial conditions). This

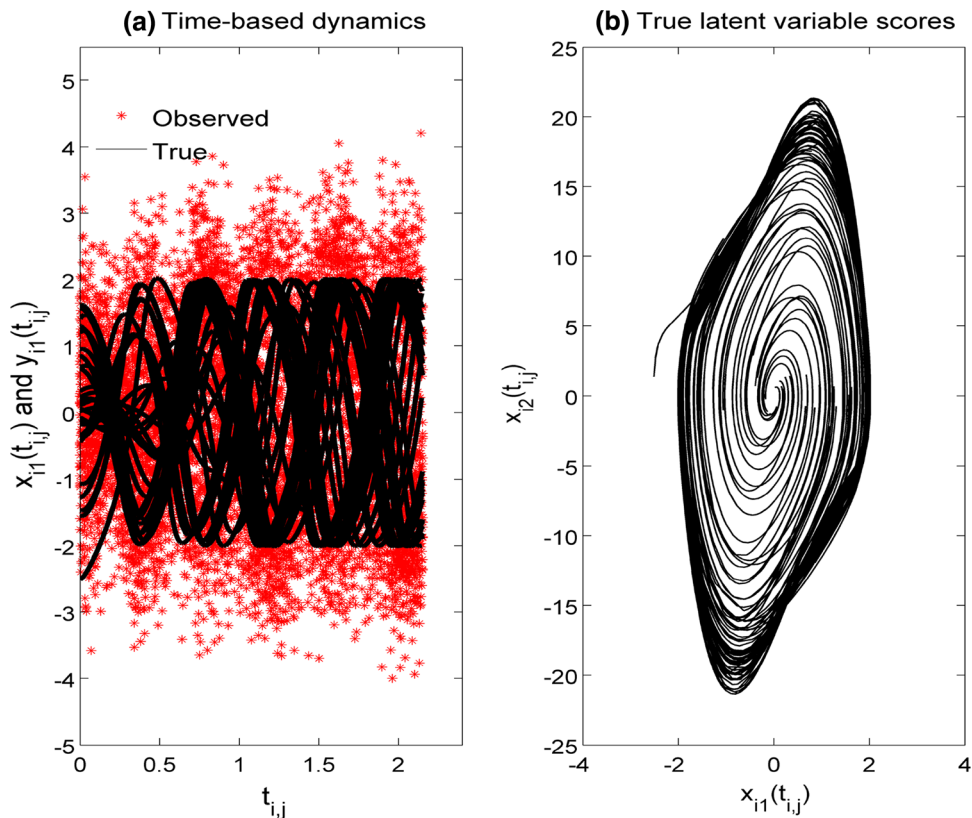


FIGURE 2.

a Plot of the true $x_{i1}(t_{i,j})$ and noisy observations, $y_{i1}(t_{i,j})$, of 50 randomly selected subjects from the Van der Pol model over time; **b** plot of two of the latent variables, $x_{i1}(t_{i,j})$ and its first derivatives, $x_{i2}(t_{i,j})$, over all time points.

yielded a total of four true–fitted initial condition specifications, with fixed–fixed, fixed–random, random–fixed, and random–random initial conditions. Among the conditions with fixed true initial condition specification, the fixed–fixed configuration was expected to yield slightly better performance (e.g., with fewer numerical problems) than the fixed–random condition, which is also able to capture the fixed initial conditions as a special case of our proposed modeling framework.

Among the conditions with random true initial condition specification, the random–random configuration was expected to yield substantially better performance than the random–fixed configuration in which μ_{x1} and μ_{x2} (which had a true value of 0) were fixed at the incorrect constant value of 1.0, with no interindividual differences therein. Note that this configuration is worth considering because this is a strategy often adopted by uninformed users of ODE solvers who, in the absence of knowledge concerning the true initial conditions, resort to fixing the initial conditions at arbitrary constant values. Results across the different true/fitted initial condition specifications also help provide insights into the sensitivity of modeling results to misspecification in the true initial conditions.

Simulated data from 50 randomly selected subjects are plotted in Figures 2a, b. In this particular parameter range, the Van der Pol oscillator model is expected to yield limit cycle behavior (Strogatz, 1994), or in other words, ongoing, isolated oscillations that either attract or repel neighboring trajectories (i.e., trajectories started out with similar values would either spiral toward or away from the cycle). Unlike cyclic oscillations that arise in linear dynamic systems

where the amplitudes of oscillations are determined solely by initial conditions, the amplitudes of the oscillations in a Van der Pol system (or other similar systems) depend on latent variables in the system (see Strogatz, 1994, pp. 196–227, chap. 7). In this way, systems such as the Van der Pol model are appropriate for representing natural systems that exhibit self-sustained oscillations.

To summarize, we considered 2 (sample size configurations) \times 2 (true initial conditions) \times 2 (fitted initial conditions) = 8 conditions in our simulation study. Statistical properties of the point and standard error (SE) estimates over 200 Monte Carlo replications, and the corresponding average correlations between the true and estimated $b_{\zeta,i}$ across the 8 conditions, are shown in Tables 2 and 9. The root mean squared errors (RMSEs) and relative biases were used to quantify the performance of the point estimates. The empirical SE of a parameter (i.e., the standard deviation of the parameter estimates across all Monte Carlo runs) was used as the “true” standard error. As a measure of the relative performance of the SE estimates, we used the average relative deviance of a SE estimate of an estimator (denoted as RDSE, namely, the difference between the average SE estimate and the true SE over the true SE). Ninety-five percent confidence intervals were constructed for each of the 200 simulation samples by adding and subtracting $1.96 * \widehat{SE}$ in each replication to the parameter estimate from the replication. We then computed power estimates⁶ by tallying the proportion of Monte Carlo trials in which the 95% CIs did not include zero. For parameters that had a true value of zero (i.e., $\mu_1 - \mu_3$, and $\sigma_{b_{x1}}^2$, $\sigma_{b_{x2}}^2$ and $\sigma_{b_{x1,x2}}$ in the fixed–random condition), this proportion can be taken as a type I error estimate, namely, the proportion of Monte Carlo trials in which the true parameter values of zero were incorrectly concluded as statistically significantly different from zero. To minimize the effects of outliers, we screened for cases manually, as well as winsorized 5% of the most extreme estimates (i.e., replacing cases that were lower or higher than the 5th and 95th percentiles by values of the 5th and 95th percentiles, respectively) before comparing the results across conditions. The percentages of retained cases used for comparison purposes are summarized in the footnotes of Tables 2 and 3.

We first focus on elaborating results from the fixed–fixed conditions (i.e., with fixed true and fitted initial conditions), as this is the specification typically assumed in comparing ODE modeling methods. Results indicated that biases of the point estimates (as indicated by relative biases and RMSEs) and discrepancies in SE estimates in comparison with the MC SDs (as indicated by RDSEs) were relatively small. Biases in the point and SE estimates were both observed to decrease with increasing time points (see Tables 2, 3). Slightly large RMSEs, relative biases, and RDSEs were observed in the point and SE estimates of the fixed effects dynamic parameters, $\zeta_0 - \zeta_3$, and the random effect variance, $\sigma_{b_{\zeta}}^2$. This is to be expected, because a second-order ODE solver was used to approximate the trajectories generated using a fourth-order ODE solver. In other words, some approximation errors were inherently present in estimates of the latent variables, namely, $\tilde{\mathbf{x}}_i(t_{i,j})$, especially with large time intervals. These biases, in turn, would likely affect estimates of the dynamic parameters in $\theta_{f,i}$. The biases in point estimates were still within reasonable ranges given the time intervals considered in this simulation study; a larger number of time points and/or participants are likely needed to further improve properties of the point and SE estimates for all of the dynamic and random effect-related parameters.

The discrepancies in SE estimates reduced considerably from $T = 150$ to $T = 300$, with the exception of the fixed effects parameters for the two covariates, ζ_1 and ζ_2 , which did not show clear gain in accuracy and precision from $T = 150$ to 300. Thus, consistent with findings concerning power issues in growth curve modeling (Raudenbush & Liu, 2001), it may be necessary to increase the number of participants to improve the estimation properties of ζ_1 and ζ_2 and the

⁶ Because the model fitting procedures were based on the second-order Heun’s method whereas the true data were generated using a fourth-order Runge Kutta approach, the errors entailed from approximating the trajectories from the fourth-order solver by means of a second-order solver were expected to lead to some biases in the point estimates. Thus, the coverage performance of the confidence intervals as assessed, e.g., by the proportion of 95% CIs covering each true population parameter value, can be expected to deviate from the nominal coverage rate of 0.95.

TABLE 2.

Parameter estimates for the Van der Pol oscillator model with $T = 150$, true initial condition = fixed, fitted initial condition = fixed.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	$a\widehat{SE}$	MC SD	RDSE	Power/type I error
ζ_0	3.00	3.05	0.05	0.02	0.077	0.318	-0.76	1.00
ζ_1	0.50	0.48	0.02	-0.04	0.018	0.048	-0.62	1.00
ζ_2	0.50	0.49	0.01	-0.01	0.018	0.058	-0.69	1.00
μ_1	0.00	-0.00	0.00	-	0.004	0.006	-0.25	0.16
μ_2	0.00	-0.00	0.00	-	0.004	0.005	-0.11	0.12
μ_3	0.00	-0.00	0.00	-	0.004	0.007	-0.38	0.21
λ_{21}	0.70	0.70	0.00	0.00	0.003	0.004	-0.11	1.00
λ_{31}	1.20	1.20	0.00	0.00	0.003	0.005	-0.37	1.00
$\sigma_{\epsilon_1}^2$	0.50	0.50	0.00	-0.00	0.004	0.004	-0.02	1.00
$\sigma_{\epsilon_2}^2$	0.50	0.50	0.00	-0.00	0.004	0.004	0.14	1.00
$\sigma_{\epsilon_3}^2$	0.50	0.50	0.00	-0.00	0.004	0.004	-0.07	1.00
$\sigma_{b_\zeta}^2$	0.50	0.48	0.02	-0.04	0.048	0.123	-0.61	1.00

% of retained cases = 96%, correlation between true and estimated $b_{\zeta,i} = 0.87$, true θ = true value of a parameter, mean $\hat{\theta} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h$, where $\hat{\theta}_h$ = estimate of θ from the h th Monte Carlo replication, $RMSE = \sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta)^2}$, $rBias$ relative bias = $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta) / \text{true } \theta$, SE standard deviation of $\hat{\theta}$ across Monte Carlo runs, $a\widehat{SE}$ = average standard error estimate across Monte Carlo runs, $RDSE$ average relative deviance of $\widehat{SE} = (a\widehat{SE} - SE) / SE$, power/type I error = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications.

TABLE 3.

Parameter estimates for the Van der Pol oscillator model with $T = 300$, true initial condition = fixed, fitted initial condition = fixed.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	$a\widehat{SE}$	MC SD	RDSE	Power/type I error
ζ_0	3.00	2.96	0.04	-0.01	0.056	0.173	-0.68	1.00
ζ_1	0.50	0.49	0.01	-0.01	0.014	0.049	-0.72	1.00
ζ_2	0.50	0.49	0.01	-0.01	0.013	0.042	-0.68	1.00
μ_1	0.00	-0.00	0.00	-	0.003	0.004	-0.19	0.13
μ_2	0.00	-0.00	0.00	-	0.003	0.003	-0.12	0.09
μ_3	0.00	-0.00	0.00	-	0.003	0.004	-0.17	0.12
λ_{21}	0.70	0.70	0.00	0.00	0.002	0.002	0.05	1.00
λ_{31}	1.20	1.20	0.00	0.00	0.002	0.003	-0.15	1.00
$\sigma_{\epsilon_1}^2$	0.50	0.50	0.00	-0.00	0.003	0.003	0.12	1.00
$\sigma_{\epsilon_2}^2$	0.50	0.50	0.00	0.00	0.003	0.002	0.16	1.00
$\sigma_{\epsilon_3}^2$	0.50	0.50	0.00	-0.00	0.003	0.003	-0.03	1.00
$\sigma_{b_\zeta}^2$	0.50	0.51	0.01	0.02	0.050	0.083	-0.40	1.00

% of retained cases = 94%, correlation between true and estimated $b_{\zeta,i} = 0.92$, true θ = true value of a parameter, mean $\hat{\theta} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h$, where $\hat{\theta}_h$ = estimate of θ from the h th Monte Carlo replication, $RMSE = \sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta)^2}$, $rBias$ relative bias = $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta) / \text{true } \theta$, SE standard deviation of $\hat{\theta}$ across Monte Carlo runs, $a\widehat{SE}$ = average standard error estimate across Monte Carlo runs, $RDSE$ average relative deviance of $\widehat{SE} = (a\widehat{SE} - SE) / SE$, power/type I error = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications.

TABLE 4.

Parameter estimates for the Van der Pol oscillator model with $T = 150$, true initial condition = fixed, fitted initial condition = random.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	$a\widehat{SE}$	MC SD	RDSE	Power/type I error
ζ_0	3.00	3.23	0.23	0.08	0.075	0.133	-0.44	1.00
ζ_1	0.50	0.45	0.05	-0.09	0.017	0.036	-0.53	1.00
ζ_2	0.50	0.45	0.05	-0.09	0.017	0.039	-0.56	1.00
μ_{x_1}	1.00	1.00	0.00	-0.00	0.008	0.012	-0.38	1.00
μ_{x_2}	1.00	1.01	0.01	0.01	0.021	0.034	-0.39	1.00
μ_1	0.00	-0.00	0.00	-	0.004	0.004	-0.05	0.12
μ_2	0.00	-0.00	0.00	-	0.004	0.004	0.03	0.00
μ_3	0.00	-0.00	0.00	-	0.004	0.005	-0.13	0.14
λ_{21}	0.70	0.70	0.00	0.00	0.003	0.003	-0.09	1.00
λ_{31}	1.20	1.20	0.00	0.00	0.003	0.004	-0.23	1.00
$\sigma_{\epsilon_1}^2$	0.50	0.50	0.00	-0.00	0.004	0.005	-0.15	1.00
$\sigma_{\epsilon_2}^2$	0.50	0.50	0.00	-0.00	0.004	0.004	0.05	1.00
$\sigma_{\epsilon_3}^2$	0.50	0.50	0.00	-0.00	0.004	0.004	0.00	1.00
$\sigma_{b_{\zeta}}^2$	0.50	0.46	0.04	-0.08	0.046	0.134	-0.66	1.00
$\sigma_{b_{x_1}}^2$	0.00	0.01	0.01	-	0.001	0.004	-0.68	1.00
$\sigma_{b_{x_2}}^2$	0.00	0.10	0.10	-	0.103	0.022	3.69	0.00
$\sigma_{b_{x_1,x_2}}$	0.00	-0.02	0.02	-	0.010	0.006	0.54	0.26

% of retained cases = 86 %, correlation between true and estimated $b_{\zeta,i} = 0.79$, true θ = true value of a parameter, mean $\hat{\theta} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h$, where $\hat{\theta}_h$ = estimate of θ from the h th Monte Carlo replication, RMSE = $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta)^2}$, $rBias$ relative bias = $\frac{1}{H} \sum_h (\hat{\theta}_h - \text{true } \theta) / \text{true } \theta$, SE standard deviation of $\hat{\theta}$ across Monte Carlo runs, $a\widehat{SE}$ = average standard error estimate across Monte Carlo runs, $RDSE$ average relative deviance of $\widehat{SE} = (a\widehat{SE} - SE) / SE$, power/type I error = 1 - the proportion of 95 % confidence intervals (CIs) that contain 0 across the Monte Carlo replications.

initial condition parameters. In addition, power estimates were generally close to 1.00 for the sample sizes considered in this simulation study, although the type I error rates associated with the three intercept parameters, $\mu_1 - \mu_3$, whose true values were equal to zero, also appeared slightly elevated.

The performance measures just noted for the fixed–fixed conditions can be contrasted directly with the measures obtained from the fixed–random (i.e., fixed true initial conditions and random fitted initial conditions) because the same 200 sets of MC data were used to compare the performances of the two fitted initial condition specifications. Results indicated that specifying the initial conditions as conforming to a multivariate distribution with unknown mean vector and covariance matrix still led to satisfactory estimation results (see Tables 4, 5). In particular, all the measurement parameters remained unbiased and showed comparable levels of precision (in terms of MC SDs) compared to conditions with the fixed–fixed specification at equivalent sample sizes. Slight increases in biases were observed for the three fixed effects dynamic parameters, $\zeta_0 - \zeta_3$; however, higher precision (i.e., reduced MC SDs) and smaller RDSEs were obtained for almost all of the parameters. When the initial conditions were fixed at known and correctly specified values, the algorithm was able to yield close to unbiased point estimates for $\zeta_0 - \zeta_3$, but at the expense of lower precision, possibly due to the approximation errors. In contrast, when the initial conditions were freely estimated, even though greater biases were present in $\zeta_0 - \zeta_3$ and the average correlation between the true and estimated $b_{\zeta,i}$ did decrease slightly (see footnotes of Tables 2,

TABLE 5.

Parameter estimates for the Van der Pol oscillator model with $T = 300$, true initial condition = fixed, fitted initial condition = random.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	$a\widehat{SE}$	MC SD	RDSE	Power/type I error
ζ_0	3.00	3.22	0.22	0.07	0.054	0.124	-0.56	1.00
ζ_1	0.50	0.45	0.05	-0.10	0.012	0.031	-0.61	1.00
ζ_2	0.50	0.45	0.05	-0.10	0.012	0.030	-0.60	1.00
μ_{x_1}	1.00	1.00	0.00	-0.00	0.006	0.009	-0.34	1.00
μ_{x_2}	1.00	1.01	0.01	0.01	0.015	0.024	-0.39	1.00
μ_1	0.00	-0.00	0.00	-	0.003	0.003	-0.06	0.08
μ_2	0.00	-0.00	0.00	-	0.003	0.003	0.04	0.00
μ_3	0.00	-0.00	0.00	-	0.003	0.003	0.12	0.05
λ_{21}	0.70	0.70	0.00	-0.00	0.002	0.002	-0.08	1.00
λ_{31}	1.20	1.20	0.00	-0.00	0.002	0.003	-0.27	1.00
$\sigma_{\epsilon_1}^2$	0.50	0.50	0.00	-0.00	0.003	0.003	0.07	1.00
$\sigma_{\epsilon_2}^2$	0.50	0.50	0.00	0.00	0.003	0.003	0.12	1.00
$\sigma_{\epsilon_3}^2$	0.50	0.50	0.00	-0.00	0.003	0.003	-0.03	1.00
$\sigma_{b_{\zeta}}^2$	0.50	0.49	0.01	-0.02	0.049	0.096	-0.49	1.00
$\sigma_{b_{x_1}}^2$	0.00	0.01	0.01	-	0.001	0.002	-0.68	1.00
$\sigma_{b_{x_2}}^2$	0.00	0.05	0.05	-	0.103	0.011	8.62	0.00
$\sigma_{b_{x_1,x_2}}$	0.00	-0.01	0.01	-	0.005	0.003	0.37	0.41

% of retained cases = 98 %, correlation between true and estimated $b_{\zeta,i} = 0.88$, true θ = true value of a parameter, mean $\hat{\theta} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h$, where $\hat{\theta}_h$ = estimate of θ from the h th Monte Carlo replication, RMSE = $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta)^2}$, $rBias$ relative bias = $\frac{1}{H} \sum_h (\hat{\theta}_h - \text{true } \theta) / \text{true } \theta$, SE standard deviation of $\hat{\theta}$ across Monte Carlo runs, $a\widehat{SE}$ = average standard error estimate across Monte Carlo runs, $RDSE$ average relative deviance of $\widehat{SE} = (a\widehat{SE} - SE) / SE$, power/type I error = 1 - the proportion of 95 % confidence intervals (CIs) that contain 0 across the Monte Carlo replications.

3, 4, 5), the uncertainties in the initial conditions appeared to help compensate for some of these approximation errors and gave rise to higher precision and relatedly, smaller RSDEs.

The average initial condition parameters, μ_{x_1} and μ_{x_2} , were correctly estimated in the fixed-random condition, and the associated variance-covariance parameters (including $\sigma_{b_{x_1}}^2$, $\sigma_{b_{x_2}}^2$, and $\sigma_{b_{x_1,x_2}}$) were estimated to be close to the correct value of zero. The SE for $\sigma_{b_{x_2}}^2$ was clearly overestimated, however, and the type I error rates for all of the initial condition variance-covariance parameters, including $\sigma_{b_{x_1}}^2$, $\sigma_{b_{x_2}}^2$ and $\sigma_{b_{x_1,x_2}}$, also deviated quite substantially from the nominal value of 0.05. These findings did not arise in the random-random condition, and they may stem from the fact that in the fixed-random condition, we were performing estimation near the boundary values of the initial condition variance-covariance parameters, whose true values were equal to zero.⁷ A related consequence was that a higher percentage of cases with numerical problems arose in the fixed-random condition than in the fixed-fixed condition, although the percentages of replications that had converged to theoretically plausible values remained generally satisfactory (close to or above 90 %). As in the fixed-fixed condition, the fixed effects parameters for the two covariates, ζ_1 and ζ_2 , also did not show improvement in accuracy from $T = 150$ to 300, verifying

⁷ To ensure the positive definiteness of Σ_b , we chose to estimate the lower triangular entries of \mathbf{L} and the diagonal entries of \mathbf{D} in the $\Sigma_b = LDL^T$ decomposition, with the constraint that the diagonal elements of \mathbf{D} were positive (Anderson, 2003). These constraints might have affected the accuracy of the point and SE estimates for the initial condition variance-covariance parameters as well.

TABLE 6.

Parameter estimates for the Van der Pol oscillator model with $T = 150$, true initial condition = random, fitted initial condition = random.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	$a\widehat{SE}$	MC SD	RDSE	Power/type I error
ζ_0	3.00	2.92	0.08	-0.03	0.047	0.216	-0.78	1.00
ζ_1	0.50	0.46	0.04	-0.08	0.012	0.066	-0.81	1.00
ζ_2	0.50	0.46	0.04	-0.08	0.012	0.063	-0.80	1.00
μ_{x1}	0.00	0.00	0.00	-	0.001	0.069	-0.99	0.99
μ_{x2}	0.00	0.00	0.00	-	0.006	0.087	-0.93	0.95
μ_1	0.00	0.00	0.00	-	0.004	0.005	-0.11	0.10
μ_2	0.00	-0.00	0.00	-	0.004	0.004	-0.07	0.05
μ_3	0.00	0.00	0.00	-	0.004	0.005	-0.13	0.11
λ_{21}	0.70	0.70	0.00	-0.00	0.003	0.004	-0.22	1.00
λ_{31}	1.20	1.20	0.00	-0.00	0.003	0.006	-0.45	1.00
$\sigma_{\epsilon_1}^2$	0.50	0.50	0.00	0.00	0.004	0.008	-0.51	1.00
$\sigma_{\epsilon_2}^2$	0.50	0.50	0.00	0.00	0.004	0.005	-0.25	1.00
$\sigma_{\epsilon_3}^2$	0.50	0.50	0.00	0.00	0.004	0.011	-0.63	1.00
$\sigma_{b_\zeta}^2$	0.50	0.57	0.07	0.13	0.053	0.186	-0.71	1.00
$\sigma_{b_{x1}}^2$	1.00	1.10	0.10	0.10	0.110	0.116	-0.05	1.00
$\sigma_{b_{x2}}^2$	1.00	1.23	0.23	0.23	0.131	0.184	-0.29	1.00
$\sigma_{b_{x1,x2}}$	0.30	0.16	0.14	-0.45	0.123	0.112	0.09	0.29

% of retained cases =93 %, correlation between true and estimated $b_{\zeta,i} = 0.65$, true θ = true value of a parameter, mean $\hat{\theta} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h$, where $\hat{\theta}_h$ = estimate of θ from the h th Monte Carlo replication, RMSE = $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta)^2}$, $rBias$ relative bias = $\frac{1}{H} \sum_h (\hat{\theta}_h - \text{true } \theta) / \text{true } \theta$, SE standard deviation of $\hat{\theta}$ across Monte Carlo runs, $a\widehat{SE}$ = average standard error estimate across Monte Carlo runs, $RDSE$ average relative deviance of $\widehat{SE} = (a\widehat{SE} - SE) / SE$, power/type I error = 1 - the proportion of 95 % confidence intervals (CIs) that contain 0 across the Monte Carlo replications.

again the need to increase the number of subjects in future studies to improve the estimation properties of these parameters.

Our next set of findings concerns properties of the SAEM algorithm when the initial conditions were random, but paired with fitted initial conditions that were either correctly or incorrectly specified. Statistical properties of the point and SE estimates are summarized in Tables 6, 7, 8, and 9. To aid interpretation, we grouped the parameters into four major types and aggregated the performance measures by parameter type. Graphical summary of the RMSEs and RDSEs by initial condition specification and parameter type is shown in Figure 3a-d. The four parameter types considered were (1) fixed effects dynamic parameters, including $\zeta_0, \zeta_1, \zeta_2, \mu_{x1}$, and μ_{x2} ; (2) measurement parameters, including the intercept parameters $\mu_1-\mu_3$ and factor loadings λ_{21} and λ_{31} ; (3) the random effect variance for the amplification parameter, $\sigma_{b_\zeta}^2$, and (4) random effect variance and covariances for the initial conditions, including $\sigma_{b_{x1}}^2, \sigma_{b_{x2}}^2$, and $\sigma_{b_{x1,x2}}$.

Several key results can be noted from the tables with full results for the conditions with random true initial condition specification. First, specifying the mean and variance-covariance parameters of the initial condition distribution as part of the parameters in β and Σ_b led to reasonable point and SE estimates, as well estimates of b_i . In contrast, misspecifying the random initial conditions as fixed, and with μ_{x1} and μ_{x2} fixed at incorrect values, led to high biases in the point estimates of $\zeta_0-\zeta_3$, the factor loadings, and all the variance parameters. Particularly high biases were observed in the point and SE estimates for the random effect variance parameter, $\sigma_{b_\zeta}^2$. In fact, the RMSEs

TABLE 7.

Parameter estimates for the Van der Pol oscillator model with $T = 300$, true initial condition = random, fitted initial condition = random.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	$a\widehat{SE}$	MC SD	RDSE	Power/type I error
ζ_0	3.00	3.03	0.03	0.01	0.032	0.210	-0.85	1.00
ζ_1	0.50	0.44	0.06	-0.12	0.009	0.057	-0.85	1.00
ζ_2	0.50	0.45	0.05	-0.10	0.009	0.065	-0.87	1.00
μ_{x_1}	0.00	-0.00	0.00	-	0.001	0.068	-0.99	0.98
μ_{x_2}	0.00	-0.01	0.01	-	0.004	0.076	-0.95	0.92
μ_1	0.00	-0.00	0.00	-	0.003	0.003	-0.09	0.08
μ_2	0.00	-0.00	0.00	-	0.003	0.003	0.00	0.04
μ_3	0.00	-0.00	0.00	-	0.003	0.003	0.00	0.05
λ_{21}	0.70	0.70	0.00	-0.00	0.002	0.005	-0.58	1.00
λ_{31}	1.20	1.20	0.00	-0.00	0.002	0.009	-0.76	1.00
$\sigma_{\epsilon_1}^2$	0.50	0.50	0.00	0.00	0.003	0.005	-0.41	1.00
$\sigma_{\epsilon_2}^2$	0.50	0.50	0.00	0.00	0.003	0.004	-0.17	1.00
$\sigma_{\epsilon_3}^2$	0.50	0.50	0.00	0.00	0.003	0.006	-0.55	1.00
$\sigma_{b_{\zeta}}^2$	0.50	0.58	0.08	0.17	0.055	0.137	-0.60	1.00
$\sigma_{b_{x_1}}^2$	1.00	1.10	0.10	0.10	0.110	0.113	-0.03	1.00
$\sigma_{b_{x_2}}^2$	1.00	1.17	0.17	0.17	0.130	0.169	-0.23	1.00
$\sigma_{b_{x_1,x_2}}$	0.30	0.20	0.10	-0.34	0.117	0.108	0.08	0.45

% of retained cases =95%; correlation between true and estimated $b_{\zeta,i} = 0.78$, true θ = true value of a parameter, mean $\hat{\theta} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h$, where $\hat{\theta}_h$ = estimate of θ from the h th Monte Carlo replication, RMSE = $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta)^2}$, $rBias$ relative bias = $\frac{1}{H} \sum_h (\hat{\theta}_h - \text{true } \theta) / \text{true } \theta$, SE standard deviation of $\hat{\theta}$ across Monte Carlo runs, $a\widehat{SE}$ = average standard error estimate across Monte Carlo runs, $RDSE$ average relative deviance of $\widehat{SE} = (a\widehat{SE} - SE) / SE$, power/type I error = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications.

and RDSEs for the point and SE estimates of $\sigma_{b_{\zeta}}^2$ were so high for both sample size conditions that these two values were omitted from the plots in Figure 3 to avoid skewing the graphical presentation of the other conditions. Under this particular misspecification in initial conditions, the estimates of $b_{\zeta,i}$ were completely biased and showed near-zero correlations with the true $b_{\zeta,i}$ values, regardless of sample size. These results suggested that the type of misspecification in initial condition specification considered in the present simulation study can greatly compromise the quality of the estimation results—an effect that is not necessarily circumvented by increases in sample size (i.e., the number of time points).

Second, for the random–random condition, doubling the number of time points from 150 to 300 led to increases in the accuracy (in terms of relative biases and RMSEs) and precision (in terms of RMSEs and MC SDs) of most parameters, as well as an increase in the average correlation between the true and estimated values of $b_{\zeta,i}$ from 0.65 to 0.78. Third, power estimates for all of the parameters in the R–R condition were generally high, with the exception of the covariance parameter between the interindividual differences in initial level and first derivative, $\sigma_{b_{x_1,x_2}}$. “Type I error” rates for the five parameters whose true values were equal to zero ($\mu_1 - \mu_3$, μ_{x_1} , and μ_{x_2}) remained slightly elevated as in the conditions with fixed true initial conditions. Finally, slight decrements in performance were observed from the F–R conditions to the R–R conditions across both sample size configurations in terms of the quality of the point and SE estimates. Particularly notable was the increased variability (i.e., MC SDs) of all the point estimates, and

TABLE 8.

Parameter estimates for the Van der Pol oscillator model with $T = 150$, true initial condition = random, fitted initial condition = fixed.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	$a\widehat{SE}$	MC SD	RDSE	Power/type I error
ζ_0	3.00	2.62	0.38	-0.13	0.195	0.257	-0.24	1.00
ζ_1	0.50	0.85	0.35	0.70	0.047	0.056	-0.15	1.00
ζ_2	0.50	0.84	0.34	0.68	0.047	0.062	-0.24	1.00
μ_1	0.00	0.06	0.06	-	0.010	0.011	-0.07	1.00
μ_2	0.00	0.01	0.01	-	0.007	0.007	-0.03	0.17
μ_3	0.00	0.01	0.01	-	0.010	0.010	-0.07	0.20
λ_{21}	0.70	0.09	0.61	-0.87	0.005	0.041	-0.87	1.00
λ_{31}	1.20	0.15	1.05	-0.87	0.008	0.068	-0.88	1.00
$\sigma_{\zeta_1}^2$	0.50	3.22	2.72	5.44	0.026	0.195	-0.87	1.00
$\sigma_{\zeta_2}^2$	0.50	1.28	0.78	1.57	0.011	0.020	-0.47	1.00
$\sigma_{\zeta_3}^2$	0.50	2.81	2.31	4.61	0.023	0.052	-0.56	1.00
$\sigma_{b_\zeta}^2$	0.50	15.32	14.82	29.64	1.493	0.417	2.58	1.00

% of retained cases = 98%, correlation between true and estimated $b_{\zeta,i} = 0.05$, true θ = true value of a parameter, mean $\hat{\theta} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h$, where $\hat{\theta}_h$ = estimate of θ from the h th Monte Carlo replication, RMSE = $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta)^2}$, rBias relative bias = $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta) / \text{true } \theta$, SE standard deviation of $\hat{\theta}$ across Monte Carlo runs, $a\widehat{SE}$ = average standard error estimate across Monte Carlo runs, RDSE average relative deviance of $\widehat{SE} = (a\widehat{SE} - SE) / SE$, power/type I error = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications.

TABLE 9.

Parameter estimates for the Van der Pol oscillator model with $T = 300$, true initial condition = random, fitted initial condition = fixed.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	$a\widehat{SE}$	MC SD	RDSE	Power/type I error
ζ_0	3.00	2.78	0.22	-0.07	0.156	0.335	-0.53	1.00
ζ_1	0.50	0.88	0.38	0.76	0.040	0.081	-0.51	1.00
ζ_2	0.50	0.87	0.37	0.74	0.039	0.078	-0.50	1.00
μ_1	0.00	0.09	0.09	-	0.007	0.009	-0.18	1.00
μ_2	0.00	0.01	0.01	-	0.005	0.005	-0.12	0.38
μ_3	0.00	0.01	0.01	-	0.007	0.008	-0.12	0.44
λ_{21}	0.70	0.07	0.63	-0.89	0.004	0.048	-0.92	0.99
λ_{31}	1.20	0.13	1.07	-0.89	0.006	0.081	-0.93	0.98
$\sigma_{\zeta_1}^2$	0.50	3.33	2.83	5.67	0.019	0.214	-0.91	1.00
$\sigma_{\zeta_2}^2$	0.50	1.29	0.79	1.58	0.007	0.019	-0.60	1.00
$\sigma_{\zeta_3}^2$	0.50	2.82	2.32	4.64	0.016	0.052	-0.69	1.00
$\sigma_{b_\zeta}^2$	0.50	15.69	15.19	30.39	1.521	0.474	2.21	1.00

% of retained cases = 100%, correlation between true and estimated $b_{\zeta,i} = 0.05$, true θ = true value of a parameter, mean $\hat{\theta} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h$, where $\hat{\theta}_h$ = estimate of θ from the h th Monte Carlo replication, RMSE = $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta)^2}$, rBias relative bias = $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \text{true } \theta) / \text{true } \theta$, SE standard deviation of $\hat{\theta}$ across Monte Carlo runs, $a\widehat{SE}$ = average standard error estimate across Monte Carlo runs, RDSE average relative deviance of $\widehat{SE} = (a\widehat{SE} - SE) / SE$, power/type I error = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications.

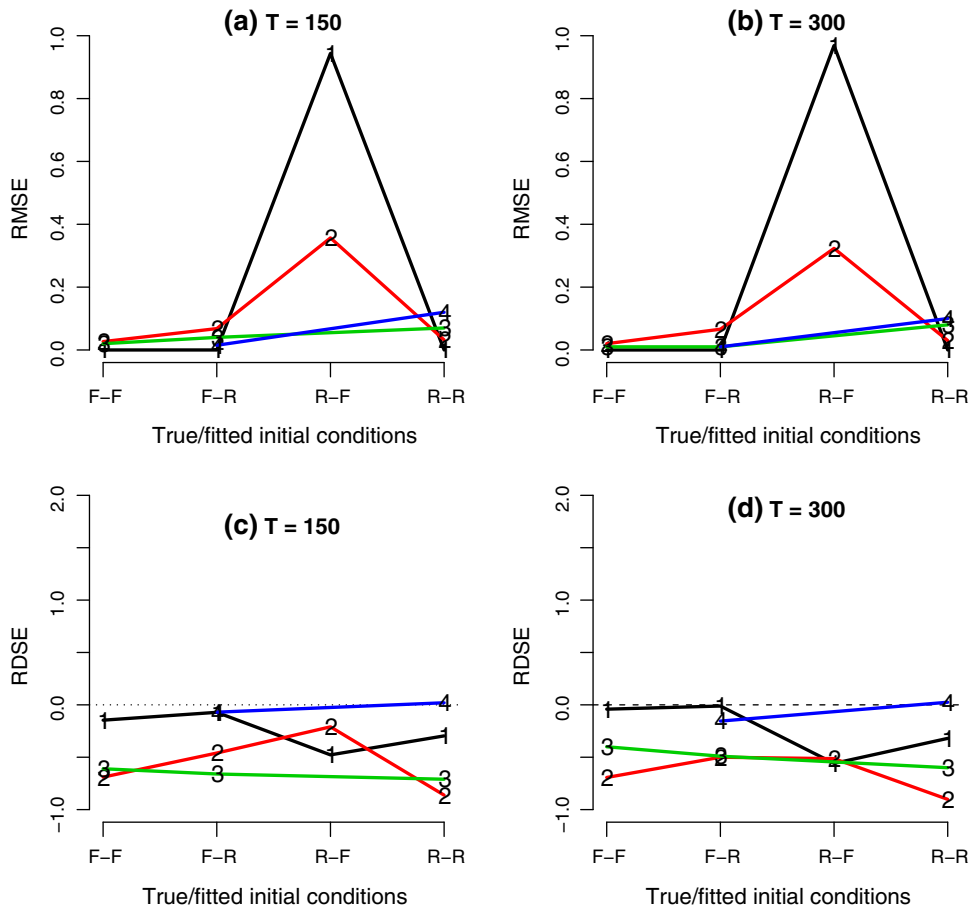


FIGURE 3.

a–d Plots of the RMSEs and RDSEs across the four true–fitted initial condition specifications and two sample size configurations. *F–F* Fixed–fixed, *F–R* fixed–random, *R–F* random–fixed, *R–R* random–random. The numbers in the plots indicate parameter type: 1 fixed effects dynamic parameters, including ζ_0 , ζ_1 , ζ_2 , μ_{x1} , and μ_{x2} , 2 measurement parameters, including the intercept parameters μ_1 – μ_3 and factor loadings λ_{21} and λ_{31} , 3 the random effect variance for the amplification parameter, $\sigma_{b_\zeta}^2$, and 4 random effect variance and covariances for the initial conditions, including $\sigma_{b_{x1}}^2$, $\sigma_{b_{x2}}^2$ and $\sigma_{b_{x1,x2}}$. To avoid skewing the graphical presentation of results from the remaining conditions, the high RMSEs and RDSEs for $\sigma_{b_\zeta}^2$ (i.e., parameter type 3) in the R–F condition were omitted from the plots.

the higher RDSEs for the fixed effects dynamic and measurement parameters (i.e., parameter types 1 and 2 in Figure 3). Despite these decrements in performance, the estimates from the R–R condition were still far better than those from the R–F conditions, whose estimation was based on the same 200 sets of MC data as the former. Even though the RDSEs for the dynamic and measurement parameters (parameter types 1 and 2) appeared reasonable for the R–F conditions, the corresponding point and $b_{\zeta,i}$ estimates were too biased to be practically useful.

4. Discussion

In the present article, we presented an SAEM algorithm for fitting linear or nonlinear dynamic models with random effects in the dynamic parameters and unknown initial conditions. Although

the illustrative and simulation examples are all nonlinear in nature, the proposed algorithm is applicable to linear ODEs as special cases. Using a modified Van der Pol oscillator model as an illustrative model, we evaluated the estimation properties of the proposed technique using an empirical example and a simulation study.

Our simulation results indicated that the proposed technique yielded satisfactory point and SE estimates for most of the parameters. Further developments are needed, however, to improve the accuracy of the SE estimates of the dynamic parameters. The problem is especially pronounced in cases involving random, as opposed to fixed initial conditions. We also demonstrated the feasibility of the proposed approach in handling situations with unknown initial conditions, either with or without interindividual differences in initial conditions. In our previous work in which different initial condition specifications were compared using linear discrete-time state-space models, the proposed approach of estimating means and variance–covariance parameters of the unknown initial condition distribution as modeling parameters was found to yield reasonable estimates for a broad array of initial condition scenarios (Losardo, 2012). Here, we extended our earlier results to the case of nonlinear continuous-time models with specific parametric assumptions on the distribution of the random effect. Possible extensions may include adding covariates as predictors of the interindividual differences in initial conditions. Other alternative approaches that may be adopted in future studies include using other parametric (e.g., exponential and Laplace) and nonparametric (e.g., Chow et al., 2011) distributions for the random effects distribution, and relaxing the linear functional form of $\theta_{f,i}$ in Eq. 1.

Slight decrements in the performance of the SAEM approach were observed when the number of random effects in the model was increased from one to three. Difficulties involved in estimating multiple random effects in dynamic models are related directly to whether the structural parameters are orthogonal to each other and whether the model is empirically identifiable with multiple random effects. In our empirical example, we fixed the oscillation frequency based on the expectation of a 24-h diurnal cycle to circumvent the issue of having insufficient complete cycles of data to identify other cycles of slower frequencies. In addition, we fixed one of the factor loadings at unity to set the metrics of the latent variables. The issue of parameter and model identifiability is, however, a much more complex problem than what we have alluded to thus far. In many dynamic models, similar differences in dynamics can be attained by incorporating random effects in more than one way. Obviously, estimation issues and difficulties arise if high correlations are present among the structural parameters. To this end, techniques aimed at evaluating model identifiability and dependencies among parameters are important to consider (Miao, Xin, Perelson, & Wu, 2011).

We adopted a Frequentist approach to parameter estimation because it offers a well-understood framework for performing hypothesis testing, and explicit criteria for assessing model convergence. However, there are still some subtle differences between the proposed SAEM approach and other standard Frequentist procedures in that the random effects in \mathbf{b}_i are estimated using MCMC procedures. In this way, the proposed procedure may be regarded as a compromise between the practical advantages of the MCMC framework in handling models of high complexity and (possibly) intractable integration, and the advantages offered by the Frequentist framework in assessing the convergence of modeling parameters and performing significance tests (e.g., test of the statistical significance of the random effects variances). In our empirical example, we highlighted the possibility of using a Wald test to determine the number of random effects to be included in a model and their corresponding covariance structure. Another possibility is to perform a score test using by-products from the SAEM under a null hypothesis model that assumes no random effect in the model (Zhu & Zhang, 2006).

Compared to approaches for fitting discrete-time dynamic models, continuous-time nonlinear ODE models pose additional challenges in deriving the ODE solutions needed to obtain latent variable estimates. In the present study, we opted to use a numerical integration method of two

orders lower in the estimation process than that used in the data generation process. The estimates appeared reasonable, but improved performance may be attained with the use of a higher-order numerical solver, particularly adaptive solvers that can handle stiff systems. Other approaches using splines (Cao, Huang, & Wu, 2012; Liang, Miao, & Wu, 2010; Ramsay et al., 2007) or Langevin sampling techniques (Stuart, Voss, & Wilberg, 2004; Hairer, Stuart, Voss, & Wiberg, 2005) to replace the use of numerical solvers are also viable alternatives.

One of the primary reasons we opted to use a Monte Carlo-based EM techniques is to circumvent the difficulties involved in integrating over the random effects in b_i . One alternative to the estimation procedure proposed here is to augment the latent variable vector, $\mathbf{x}(t_{i,j})$, with \mathbf{b}_i to yield $\mathbf{x}_a(t_{i,j}) = [\tilde{\mathbf{x}}_i(t_{i,j})^T \mathbf{b}_i^T]^T$ and subsequently utilize some of the newer hybrid nonlinear filtering approaches (Kulikov & Kulikova, 2014) to obtain a direct way of computing elements such as $E(\mathbf{x}_a(t_{i,j})|\mathbf{Y}; \theta)$ required in the E-step. In this way, because the random effects are now part of the latent variable vector, no integration over $p(\mathbf{b}_i|\mathbf{Y}; \theta)$ is needed, and the E-step is thus greatly simplified. An optimization technique of choice can then be used in the M-step to update the parameter estimates. Thus, unlike our proposed approach in which the computation of the latent variable estimates is separated from the sampling of b_i , sampling of the latent variable and b_i estimates is obtained jointly in this alternative approach. The performance of these two approaches should be compared in future studies under conditions with different modeling complexities (e.g., with linear and nonlinear ODEs and SDEs), as well as different numbers of random effects.

Whereas nonlinear dynamic models open up myriad new possibilities for evaluating more complex models, many methodological issues remain unresolved and have to be handled with caution. Parallel to the increase in model complexity are, of course, new challenges in deriving appropriate model fit indices and diagnostic measures. Sensitivity of the modeling results to parameter starting values, initial condition specification, choices of ODE solvers, integration time steps, and the number of random effects parameters are all important issues that warrant further attention. In addition, we made the assumption that all individuals were characterized by the same set of ODE functions, with the only source of between-individual differences residing in the individual-specific dynamic parameters in $\theta_{f,i}$. This assumption may not be tenable in all applications and across all variables (Molenaar, 2004). Alternative formulations that combine exploratory procedures for identifying individual differences with a confirmatory framework that assumes some levels of homogeneity in the change functions may be a promising alternative (e.g., Gates & Molenaar, 2012).

Our empirical dataset was not designed with the goal to evaluate complex nonlinear dynamics and is therefore characterized by several limitations often encountered in psychological datasets. While the sample size configurations considered in our simulation study ($n = 200$, $T = 150$ or 300) may seem high to many social and behavioral scientists, particularly psychologists, multiple-subject data of such time lengths are no longer an insurmountable goal. With the advent of ecological momentary assessment designs and new technological developments for collecting data in near real-time, intensive repeated measures data have become increasingly prevalent in psychology and other related disciplines. Furthermore, the random effects framework presented in this article constitutes one possible way of pooling together information from multiple subjects, all of which may have time series data of finite lengths.

We outline some other design-related issues here in hopes of offering some suggestions to researchers interested in collecting intensive repeated measures studies for dynamic modeling purposes. First, if researchers wish to capture sustained oscillations (e.g., circadian rhythms) in a construct of interest, it is generally recommended to collect data that span at least two (or preferably more) complete cycles. This is, of course, in addition to the requirements of having sufficient time points and participants to attain reasonable estimation properties, and sampling at least twice as fast as the frequency of interest. Second, smaller time intervals (and hence integration time steps)

typically help improve the numerical stability of the estimation algorithm. In practice, the time scale used in model fitting can be rescaled (e.g., from milliseconds to hours) to yield smaller time intervals. However, if the time steps are too irregularly spaced, simple rescaling per se does not alleviate the problem. For instance, in our empirical data, the participants' integration time steps ranged from 0.001 to 0.48. A simple rescaling may yield unnecessarily high computational costs at the smaller time steps while numerical instability may still arise at the larger time steps. Thus, researchers may want to consider the timing of change of the process of interest in selecting appropriate measurement intervals for a study. Finally, researchers should consider design-related enhancements to ensure that they sample individuals sufficiently around the times when complex, nonlinear changes are expected to unfold rapidly. Otherwise, some of the critical dynamics of a system may still be bypassed even with a very large number of time points. Other design-related issues implicated in the study of dynamic processes have also been discussed by Wu (2005).

All the models presented in this article were fitted using our own scripts written in MATLAB. Other statistical programs that can handle matrix operations, such as R (2009) and SAS/IML (2008), may also be used. Because the estimation algorithm was written in MATLAB and not using a primary programming language such as Fortran or C++, the computational time is relatively high. For instance, each replication in our simulation study may take 5–9 h—depending on the sample size and the number of random effects in the model—using a single core of an Intel Xeon Processor X5560 with less than 2 GB of memory. Ultimately, the high computational time is related directly to the number of time points available from each participant, because the numerical solution for each participant at each time point has to be derived sequentially based on the numerical solution for the participant at the previous time point. However, we still note that if the algorithm is reprogrammed in a more efficient primary programming language, the amount of computational time should be shorter compared to the typical computational time needed to estimate a model of comparable complexity using a fully Bayesian approach. This is because in the latter, there is no explicit criterion to guide convergence decisions; a large number of burn-in iterations has to be included before any inferences can be made utilizing the posterior distributions. In contrast, because the SAEM does not use a fully Bayesian approach, our experience has suggested that the number of burn-in iterations required to complete the two computational stages of the SAEM is substantially less than that required to complete the computations in a fully Bayesian approach.

As a discipline, psychology has generally seemed reluctant to embrace the promises and relatedly, some of the data collection/analytic challenges brought on by dynamic systems modeling techniques. Methodological difficulties and lack of understanding concerning the strengths and limitations of nonlinear systems have fostered a deeply rooted bias against nonlinear models—an unfortunate limit imposed by methodology on theory. Undoubtedly, several methodological issues remain in evaluating nonlinear dynamic models with random effects and related variations. Nonetheless, developing and evaluating estimation techniques that allow a direct mapping between more complex mathematical models of change and empirical measurements are among the first steps toward revoking the belief that nonlinear dynamic models are but a theoretical metaphor in social and behavioral research (see e.g., Kincanon & Powel, 1995). Our hope is that the work presented here can help inspire more research into alternative methods that are suited for studying change—including both linear and nonlinear dynamic models.

Acknowledgments

Funding for this study was provided by NSF Grant BCS-0826844, NIH Grants RR025747-01, P01CA142538-01, MH086633, EB005149-01, AG033387, and R01GM105004.

Appendix 1: Score Vector and Information Matrix of the Complete-Data Loglikelihood Function

The elements in $s_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})$ and $\mathbf{I}_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z})$, namely, the score vector and information matrix of the complete-data loglikelihood function, are computed as

$$s_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z}) = \frac{\partial L(\mathbf{Z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \begin{bmatrix} \left(\sum_{j=1}^T \frac{\partial L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \right) \\ \left(\sum_{j=1}^T \frac{\partial L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\mu}} \right) \\ \left(\sum_{j=1}^T \frac{\partial L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\Lambda}} \right) \\ \left(\sum_{j=1}^T \frac{\partial L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\epsilon}} \right) \\ \left(\frac{\partial L_i(\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\mathbf{b}}} \right) \end{bmatrix}, \quad (17)$$

$$\mathbf{I}_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z}) = -\frac{\partial^2 L(\mathbf{Z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = -\sum_{i=1}^n \text{Diag} \begin{bmatrix} \sum_{j=1}^T \frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \\ \sum_{j=1}^T \frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\mu} \partial \boldsymbol{\theta}_{\mu}^T} \\ \sum_{j=1}^T \frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\Lambda} \partial \boldsymbol{\theta}_{\Lambda}^T} \\ \sum_{j=1}^T \frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\epsilon} \partial \boldsymbol{\theta}_{\epsilon}^T} \\ \frac{\partial^2 L_i(\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\mathbf{b}} \partial \boldsymbol{\theta}_{\mathbf{b}}^T} \end{bmatrix}, \quad (18)$$

where $\text{Diag}(\cdot)$ denotes a block diagonal matrix formed by stacking the appropriate second partial derivative matrices in its diagonal section and zero matrices in its off-diagonal sections.

Using Heun's method, with $\tilde{\mathbf{x}}_i(t_{i,j})$ as defined in Eq. (4) and $\mathbf{z}_i(t_{i,j}) = [\mathbf{y}_i(t_{i,j}) - \boldsymbol{\mu} - \boldsymbol{\Lambda} \tilde{\mathbf{x}}_i(t_{i,j})]$, first-order partial derivative elements of the complete-data loglikelihood are given by

$$\begin{aligned} \frac{\partial L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= -\frac{1}{2} \left\{ \frac{\partial \boldsymbol{\theta}_{f,i}}{\partial \boldsymbol{\beta}} \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})}{\partial \boldsymbol{\theta}_{f,i}} \frac{\partial \mathbf{z}_i(t_{i,j})}{\partial \tilde{\mathbf{x}}_i(t_{i,j})} \frac{\partial \mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{z}_i(t_{i,j})}{\partial \mathbf{z}_i(t_{i,j})} \right\} \\ &= \mathbf{H}_i^T \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T}{\partial \boldsymbol{\theta}_{f,i}} \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{z}_i(t_{i,j}), \\ \frac{\partial L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\mu}} &= -\frac{1}{2} \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_{\mu}} \frac{\partial \mathbf{z}_i(t_{i,j})}{\partial \boldsymbol{\mu}} \frac{\partial \mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{z}_i(t_{i,j})}{\mathbf{z}_i(t_{i,j})} \right\} \\ &= \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_{\mu}} \boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{z}_i(t_{i,j}), \\ \frac{\partial L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\Lambda}} &= -\frac{1}{2} \left\{ \frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}_{\Lambda}} \frac{\partial \mathbf{z}_i(t_{i,j})}{\partial \text{vec}(\boldsymbol{\Lambda})} \frac{\partial \mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{z}_i(t_{i,j})}{\partial \mathbf{z}_i(t_{i,j})} \right\} \\ &= \frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}_{\Lambda}} \text{vec} \left[\boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{z}_i(t_{i,j}) \tilde{\mathbf{x}}_i(t_{i,j})^T \right], \\ \frac{\partial L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{\epsilon}} &= -\frac{1}{2} \frac{\partial \text{vec}(\boldsymbol{\Sigma}_{\epsilon})}{\partial \boldsymbol{\theta}_{\epsilon}} \left\{ \frac{\partial \mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{z}_i(t_{i,j})}{\partial \text{vec}(\boldsymbol{\Sigma}_{\epsilon})} + \frac{\partial \log |\boldsymbol{\Sigma}_{\epsilon}|}{\partial \text{vec}(\boldsymbol{\Sigma}_{\epsilon})} \right\} \\ &= \frac{1}{2} \frac{\partial \text{vec}(\boldsymbol{\Sigma}_{\epsilon})}{\partial \boldsymbol{\theta}_{\epsilon}} \left\{ [\boldsymbol{\Sigma}_{\epsilon}^{-1} \otimes \boldsymbol{\Sigma}_{\epsilon}^{-1}] \text{vec}[\mathbf{z}_i(t_{i,j}) \mathbf{z}_i(t_{i,j})^T - \boldsymbol{\Sigma}_{\epsilon}] \right\}, \end{aligned}$$

$$\begin{aligned}\frac{\partial L_i(\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_b} &= -\frac{1}{2} \frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b} \left\{ \frac{\partial \mathbf{b}_i^T \boldsymbol{\Sigma}_b^{-1} \partial \mathbf{b}_i}{\partial \text{vec}(\boldsymbol{\Sigma}_b)} + \frac{\partial \log |\boldsymbol{\Sigma}_b|}{\partial \text{vec}(\boldsymbol{\Sigma}_b)} \right\} \\ &= \frac{1}{2} \frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b} \left\{ [\boldsymbol{\Sigma}_b^{-1} \otimes \boldsymbol{\Sigma}_b^{-1}] \text{vec}(\mathbf{b}_i \mathbf{b}_i^T - \boldsymbol{\Sigma}_b) \right\},\end{aligned}\quad (19)$$

where the $\text{vec}(\mathbf{W})$ operator stacks the columns of the $m \times n$ matrix \mathbf{W} into an mn -dimensional column vector and $\frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T}{\partial \boldsymbol{\theta}_{f,i}}$ is dictated by the dynamic model under consideration. Terms such as $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_\mu}$, $\frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}_\Lambda}$, $\frac{\partial \text{vec}(\boldsymbol{\Sigma}_\epsilon)}{\partial \boldsymbol{\theta}_\epsilon}$, $\frac{\partial \text{vec}(\boldsymbol{\Sigma}_\eta)}{\partial \boldsymbol{\theta}_\eta}$, and $\frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b}$ also depend on the model specification adopted in a particular application. Cases where some elements of $\boldsymbol{\Lambda}$, $\boldsymbol{\Sigma}_\epsilon$, $\boldsymbol{\Sigma}_\eta$, and $\boldsymbol{\Sigma}_b$ are fixed at known values can be readily accommodated through appropriate specification of these matrices of partial derivatives.

Second-order partial derivative elements of the complete-data loglikelihood function are computed as

$$\begin{aligned}\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= \left(\mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{\Lambda}^T \otimes \mathbf{I}_{p_\beta} \right) \frac{\partial \text{vec}}{\partial \boldsymbol{\beta}^T} \left\{ \mathbf{H}_i^T \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T}{\partial \boldsymbol{\theta}_{f,i}} \right\} \\ &\quad + \left(\mathbf{H}_i^T \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T}{\partial \boldsymbol{\theta}_{f,i}} \right) \frac{\partial \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{z}_i(t_{i,j})}{\partial \mathbf{z}_i(t_{i,j})^T} \frac{\partial \mathbf{z}_i(t_{i,j})}{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T} \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})}{\partial \boldsymbol{\theta}_{f,i}^T} \frac{\partial \boldsymbol{\theta}_{f,i}}{\partial \boldsymbol{\beta}^T} \\ &= \left(\mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{\Lambda} \otimes \mathbf{I}_{p_\beta} \right) (\mathbf{I}_{n_x} \otimes \mathbf{H}_i^T) \frac{\partial \text{vec}}{\partial \boldsymbol{\theta}_{f,i}^T} \left\{ \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T}{\partial \boldsymbol{\theta}_{f,i}} \right\} \mathbf{H}_i \\ &\quad - \mathbf{H}_i^T \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T}{\partial \boldsymbol{\theta}_{f,i}} \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{\Lambda} \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})}{\partial \boldsymbol{\theta}_{f,i}^T} \mathbf{H}_i, \\ \frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\mu \partial \boldsymbol{\theta}_\mu^T} &= \left(\mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_\epsilon^{-1} \otimes \mathbf{I}_{p_\mu} \right) \frac{\partial \text{vec}}{\partial \boldsymbol{\theta}_\mu^T} \left\{ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_\mu} \right\} - \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_\mu} \boldsymbol{\Sigma}_\epsilon^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_\mu^T}, \\ \frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\Lambda \partial \boldsymbol{\theta}_\Lambda^T} &= \left\{ \text{vec}[\boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{z}_i(t_{i,j}) \tilde{\mathbf{x}}_i(t_{i,j})^T] \otimes \mathbf{I}_{p_\Lambda} \right\} \frac{\partial \text{vec}}{\partial \boldsymbol{\theta}_\Lambda^T} \left\{ \frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}_\Lambda} \right\} \\ &\quad - \frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}_\Lambda} [\tilde{\mathbf{x}}_i(t_{i,j}) \tilde{\mathbf{x}}_i(t_{i,j})^T \otimes \boldsymbol{\Sigma}_\epsilon^{-1}] \frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}_\Lambda^T}, \\ \frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\epsilon \partial \boldsymbol{\theta}_\epsilon^T} &= \frac{1}{2} \left[\text{vec} \left(\boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{z}_i(t_{i,j}) \mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_\epsilon^{-1} \right)^T \otimes \mathbf{I}_{p_\epsilon} \right] \frac{\partial \text{vec}}{\partial \boldsymbol{\theta}_\epsilon^T} \left\{ \frac{\partial \text{vec}(\boldsymbol{\Sigma}_\epsilon)}{\partial \boldsymbol{\theta}_\epsilon} \right\} \\ &\quad - \frac{\partial \text{vec}(\boldsymbol{\Sigma}_\epsilon)}{\partial \boldsymbol{\theta}_\epsilon} [\boldsymbol{\Sigma}_\epsilon^{-1} \otimes \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{z}_i(t_{i,j}) \mathbf{z}_i(t_{i,j})^T \boldsymbol{\Sigma}_\epsilon^{-1}] \frac{\partial \text{vec}(\boldsymbol{\Sigma}_\epsilon)}{\partial \boldsymbol{\theta}_\epsilon^T} \\ &\quad - \frac{1}{2} [\text{vec}(\boldsymbol{\Sigma}_\epsilon^{-1})^T \otimes \mathbf{I}_{p_\epsilon}] \frac{\partial \text{vec}}{\partial \boldsymbol{\theta}_\epsilon^T} \left\{ \frac{\partial \text{vec}(\boldsymbol{\Sigma}_\epsilon)}{\partial \boldsymbol{\theta}_\epsilon} \right\} \\ &\quad + \frac{1}{2} \frac{\partial \text{vec}(\boldsymbol{\Sigma}_\epsilon)}{\partial \boldsymbol{\theta}_\epsilon} (\boldsymbol{\Sigma}_\epsilon^{-1} \otimes \boldsymbol{\Sigma}_\epsilon^{-1}) \frac{\partial \text{vec}(\boldsymbol{\Sigma}_\epsilon)}{\partial \boldsymbol{\theta}_\epsilon^T},\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_i(\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}_b^T} &= \frac{1}{2} \left[\text{vec} \left(\boldsymbol{\Sigma}_b^{-1} \mathbf{b}_i \mathbf{b}_i^T \boldsymbol{\Sigma}_b^{-1} \right)^T \otimes \mathbf{I}_{p_b} \right] \frac{\partial \text{vec} \left\{ \frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b} \right\}}{\partial \boldsymbol{\theta}_b^T} \\
&\quad - \frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b} \left[\boldsymbol{\Sigma}_b^{-1} \otimes \boldsymbol{\Sigma}_b^{-1} \mathbf{b}_i \mathbf{b}_i^T \boldsymbol{\Sigma}_b^{-1} \right] \frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b^T} \\
&\quad - \frac{1}{2} \left[\text{vec}(\boldsymbol{\Sigma}_b^{-1})^T \otimes \mathbf{I}_{p_b} \right] \frac{\partial \text{vec} \left\{ \frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b} \right\}}{\partial \boldsymbol{\theta}_b^T} \\
&\quad + \frac{1}{2} \frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b} (\boldsymbol{\Sigma}_b^{-1} \otimes \boldsymbol{\Sigma}_b^{-1}) \frac{\partial \text{vec}(\boldsymbol{\Sigma}_b)}{\partial \boldsymbol{\theta}_b^T}, \\
\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\mu \partial \boldsymbol{\theta}_\Lambda^T} &= - \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_\mu} \boldsymbol{\Sigma}_\epsilon^{-1} [\tilde{\mathbf{x}}_i(t_{i,j})^T \otimes \mathbf{I}_{n_y}] \frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}_\Lambda^T}, \\
\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\mu \partial \boldsymbol{\beta}^T} &= - \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_\mu} \boldsymbol{\Sigma}_\epsilon^{-1} \frac{\partial z_i(t_{i,j})}{\partial \tilde{\mathbf{x}}_i(t_{i,j})^T} \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})}{\partial \boldsymbol{\theta}_{f,i}^T} \frac{\partial \boldsymbol{\beta}_{f,i}}{\partial \boldsymbol{\beta}^T}, \\
&= - \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_\mu} \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{\Lambda} \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})}{\partial \boldsymbol{\theta}_{f,i}^T} \mathbf{H}_i, \\
\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\Lambda \partial \boldsymbol{\beta}^T} &= \frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}_\Lambda} \left\{ (\mathbf{I}_{n_x} \otimes \boldsymbol{\Sigma}_\epsilon^{-1}) \left[- (\tilde{\mathbf{x}}_i(t_{i,j})^T \otimes \mathbf{I}_{n_y}) \boldsymbol{\Lambda} \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})}{\partial \boldsymbol{\theta}_{f,i}^T} \mathbf{H}_i \right. \right. \\
&\quad \left. \left. + (\mathbf{I}_{n_x} \otimes \mathbf{z}_i(t_{i,j})) \frac{\partial \tilde{\mathbf{x}}_i(t_{i,j})}{\partial \boldsymbol{\theta}_{f,i}^T} \mathbf{H}_i \right] \right\}. \tag{20}
\end{aligned}$$

Other second-order derivative elements are equal to null matrices, including $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}_{\theta_\mu}^T}$, $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}_{\theta_\epsilon}^T}$, $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}_b^T}$, $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\mu \partial \boldsymbol{\theta}_b^T}$, $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\Lambda \partial \boldsymbol{\theta}_b^T}$ and $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\epsilon \partial \boldsymbol{\theta}_b^T}$. Under the assumption that the model is correctly specified, the elements in $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\Lambda \partial \boldsymbol{\theta}_\epsilon^T}$, $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\mu \partial \boldsymbol{\theta}_\epsilon^T}$ and $\frac{\partial^2 L_{i,j}(\mathbf{Y}|\mathbf{b}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_\beta \partial \boldsymbol{\theta}_\epsilon^T}$ are close to zeros at the MLEs of the modeling parameters. These elements are thus set to null matrices in the proposed estimation algorithm to stabilize the algorithm when initial parameter estimates are far from the MLEs and are not shown here. In addition, the off-diagonal elements shown in the last three equations in (20) are non-zero even near the MLEs. However, setting all the off-diagonal blocks of the information matrix of the complete-data loglikelihood function to null matrices helps stabilize the algorithm in case this information matrix is not positive definite in the optimization process. In our preliminary simulations, we verified that setting these three matrices to null matrices, as opposed to the forms as shown in Eq. (20), actually helped reduce numerical problems in the optimization process while having negligible effects on the final point and SE estimates because we are not using this matrix directly as the Fisher information matrix to derive the final SE estimates. We thus proceeded to setting all the off-diagonal elements, including the last three matrices shown in Eq. (20), to null matrices.

Appendix 2: Sampling From $p(\mathbf{b}|\mathbf{Y}; \boldsymbol{\theta}^{(m-1)})$

The superscript $(m-1)$ of $\boldsymbol{\theta}$ is temporarily suppressed for notational simplicity. It can be shown that $p(\mathbf{b}|\mathbf{Y}; \boldsymbol{\theta}) = \prod_{i=1}^n p(\mathbf{b}_i|\mathbf{Y}; \boldsymbol{\theta})$, where $p(\mathbf{b}_i|\mathbf{Y}; \boldsymbol{\theta})$ is non-standard and cannot be sampled directly. Specifically, $p(\mathbf{b}_i|\mathbf{Y}; \boldsymbol{\theta}) \propto p(\mathbf{b}_i; \boldsymbol{\theta}_b) p(\mathbf{Y}_i|\mathbf{b}_i; \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Lambda, \boldsymbol{\theta}_\epsilon, \boldsymbol{\beta})$, in which $p(\mathbf{Y}_i|\mathbf{b}_i; \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Lambda, \boldsymbol{\theta}_\epsilon, \boldsymbol{\beta})$ is given by

$$p(\mathbf{Y}_i | \mathbf{b}_i; \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Lambda, \boldsymbol{\theta}_\epsilon, \boldsymbol{\beta}) = \prod_{j=1}^T p(y_i(t_{i,j}) | \mathbf{b}_i; \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Lambda, \boldsymbol{\theta}_\epsilon, \boldsymbol{\beta}), \quad (21)$$

where $p(y_i(t_{i,j}) | \mathbf{b}_i; \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Lambda, \boldsymbol{\theta}_\epsilon, \boldsymbol{\beta})$ is a multivariate normal density function with mean $\boldsymbol{\mu} + \boldsymbol{\Lambda} \tilde{\mathbf{x}}_i(t_{i,j})$ and covariance matrix $\boldsymbol{\Sigma}_\epsilon$. As \mathbf{b}_i is involved in the nonlinear $f(\cdot)$ in $p(y_i(t_{i,j}) | \mathbf{b}_i; \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Lambda, \boldsymbol{\theta}_\epsilon, \boldsymbol{\beta})$, $p(\mathbf{b}_i | \mathbf{Y}; \boldsymbol{\theta})$ is usually non-standard. To sample from $p(\mathbf{b}_i | \mathbf{Y}; \boldsymbol{\theta})$, we adopt a Metropolis-Hastings (MH) algorithm as follows. At the m th iteration with current values in $\mathbf{b}_i^{(m)}$, a new candidate \mathbf{b}_i is generated from a proposal distribution, chosen to be the normal distribution $N(\mathbf{b}_i^{(m)}, \sigma_b^2 \boldsymbol{\Omega}_{bi})$, where σ_b^2 is a scaling constant, $\boldsymbol{\Omega}_{bi} = (\boldsymbol{\Sigma}_b^{-1} + \sum_{j=1}^T \mathbf{D}_{bit}^T \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{D}_{bit})^{-1}$, $\mathbf{D}_{bit} = \partial \tilde{\mathbf{x}}_i(t_{i,j}) / \partial \mathbf{b}_i^T |_{\mathbf{b}_i = \mathbf{b}_i^*}$, and \mathbf{b}_i^* is a fixed value with high $p(\mathbf{b}_i^* | \mathbf{Y}; \boldsymbol{\theta})$. One possibility is to use the mean of $p(\mathbf{b}_i; \boldsymbol{\theta}_b)$ as \mathbf{b}_i^* , which we have found to lead to good performance. The new \mathbf{b}_i is accepted with probability

$$\min \left\{ 1, \frac{p(\mathbf{b}_i; \boldsymbol{\theta}_b) \prod_{j=1}^T p(y_i(t_{i,j}) | \mathbf{b}_i; \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Lambda, \boldsymbol{\theta}_\epsilon, \boldsymbol{\beta})}{p(\mathbf{b}_i^{(m)}; \boldsymbol{\theta}_b) \prod_{j=1}^T p(y_i(t_{i,j}) | \mathbf{b}_i^{(m)}; \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Lambda, \boldsymbol{\theta}_\epsilon, \boldsymbol{\beta})} \right\}, \quad (22)$$

The scaling constant, σ_b^2 , can be chosen such that the average acceptance rate is approximately 0.4.

References

- Ait-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, 36(2), 906–937.
- Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). Probability and Statistics New York, NY: Wiley.
- Arminger, G. (1986). Linear stochastic differential equation models for panel data with unobserved variables. In N. Tuma (Ed.), *Sociological methodology* (pp. 187–212). San Francisco: Jossey-Bass.
- Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3–20). Madison, WI: University of Wisconsin Press.
- Beskos, A., Papaspiliopoulos, O., & Roberts, G. (2009). Monte carlo maximum likelihood estimation for discretely observed diffusion processes. *The Annals of Statistics*, 37(1), 223–245.
- Beskos, A., Papaspiliopoulos, O., Roberts, G., & Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3), 333–382.
- Boker, S. M., & Graham, J. (1998). A dynamical systems analysis of adolescent substance abuse. *Multivariate Behavioral Research*, 33, 479–507.
- Boker, S. M., & Nesselroade, J. R. (2002). A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave panel data. *Multivariate Behavioral Research*, 37, 127–160.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616.
- Brown, E. N., & Luthardt, H. (1999). Statistical model building and model criticism for human circadian data. *Journal of Biological Rhythms*, 14, 609–616.
- Brown, E. N., Luthardt, H., & Czeisler, C. A. (2000). A statistical model of the human coretemperature circadian rhythm. *American Journal of Physiology, Endocrinology and Metabolism*, 279, 669–683.
- Browne, M. W., & du Toit, H. C. (1991). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 47–68). Washington, D.C.: American Psychological Association.
- Cao, J., Huang, J. Z., & Wu, H. (2012). Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equations. *Journal of Computational and Graphical Statistics*, 21(1), 42–56. doi:10.1198/jcgs.2011.10021.
- Carels, R. A., Blumenthal, J. A., & Sherwood, A. (2000). Emotional responsivity during daily life: Relationship to psychosocial functioning and ambulatory blood pressure. *International Journal of Psychophysiology*, 36, 25–33.
- Carlin, B. P., Gelfand, A., & Smith, A. (1992). Hierarchical bayesian analysis of changepoints problems. *Applied Statistics*, 41, 389–405.
- Chow, S.-M., Ferrer, E., & Nesselroade, J. R. (2007). An unscented kalman filter approach to the estimation of nonlinear dynamical systems models. *Multivariate Behavioral Research*, 42(2), 283–321.

- Chow, S.-M., Grimm, K. J., Guillaume, F., Dolan, C. V., & McArdle, J. J. (2013). Regime switching bivariate dual change score model. *Multivariate Behavioral Research*, *48*(4), 463–502.
- Chow, S.-M., Ho, M.-H. R., Hamaker, E. J., & Dolan, C. V. (2010). Equivalences and differences between structural equation and state-space modeling frameworks. *Structural Equation Modeling*, *17*(303–332).
- Chow, S.-M., & Nesselroade, J. R. (2004). General slowing or decreased inhibition? Mathematical models of age differences in cognitive functioning. *Journals of Gerontology Series B—Psychological Sciences & Social Sciences*, *59B*(3), 101–109.
- Chow, S.-M., Tang, N., Yuan, Y., Song, X., & Zhu, H. (2011). Bayesian estimation of semiparametric dynamic latent variable models using the Dirichlet process prior. *British Journal of Mathematical and Statistical Psychology*, *64*(1), 69–106.
- Chow, S.-M., & Zhang, G. (2013). Nonlinear regime-switching state-space (RSSS) models. *Psychometrika: Application Reviews and Case Studies*, *78*(4), 740–768.
- Cronbach, L. J., & Furby, L. (1970). How should we measure “change”—or should we? *Psychological Bulletin*, *74*(1), 68–80.
- Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods*, *7*(1), 41–6.
- Dembo, A., & Zeitouni, O. (1986). Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Processes and Their Applications*, *23*, 91–113.
- 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), 1–38.
- Diebolt, J., & Celeux, G. (1993). Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions. *Communications in Statistics B—Stochastic Models*, *9*(4), 599–613.
- Donnet, S., & Samson, A. (2007). Estimation of parameters in incomplete data models defined by dynamical systems. *Journal of Statistical Planning and Inference*, *137*, 2815–2831.
- Du Toit, S. H. C., & Browne, M. W. (2001). The covariance structure of a vector ARMA time series. *Structural equation modeling: Present and future* (pp. 279–314). Chicago: Scientific Software International.
- Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. New York, NY: Oxford University Press.
- Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *Neuroimage*, *63*, 310–319.
- Geweke, J., & Tanizaki, H. (2001). Bayesian estimation of state-space models using the Metropolis–Hastings algorithm within Gibbs sampling. *Computational Statistics & Data Analysis*, *37*, 151–170.
- Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings-F, Radar and Signal Processing*, *140*(2), 107–113.
- Gu, M. G., & Zhu, H. T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *Journal of the Royal Statistical Society, Series B*, *63*, 339–355.
- Hairer, M., Stuart, A. M., Voss, J., & Wiberg, P. (2005). Analysis of spdes arising in path sampling. part i: The gaussian case. *Communications in Mathematical Sciences*, *3*(4), 587–603.
- Hale, J. K., & Koçak, H. (1991). *Dynamics and bifurcation*. New York, NY: Springer.
- Harris, C. W. (Ed.). (1963). *Problems in measuring change*. Madison, WI: University of Wisconsin Press.
- Harvey, A. C., & Souza, R. C. (1987). Assessing and modelling the cyclical behaviour of rainfall in northeast Brazil. *Journal of Climate and Applied Meteorology*, *26*, 1317–1322.
- Hürzeler, M., & Künsch, H. (1998). Monte carlo approximations for general state-space models. *Journal of Computational and Graphical Statistics*, *7*, 175–193.
- Jones, R. H. (1984). Fitting multivariate models to unequally spaced data. In E. Parzen (Ed.), *Time series analysis of irregularly observed data* (Vol. 25, p. 158–188). New York, NY: Springer.
- Jones, R. H. (1993). *Longitudinal data with serial correlation: A state-space approach*. Boca Raton, FL: Chapman & Hall/CRC.
- Kaplan, D., & Glass, L. (1995). *Understanding nonlinear dynamics*. New York, NY: Springer.
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.
- Kincaanon, E., & Powel, W. (1995). Chaotic analysis in psychology and psychoanalysis. *The Journal of Psychology*, *129*, 495–505.
- Kitagawa, G. (1998). A self-organizing state-space model. *Journal of the American Statistical Association*, *93*(443), 1203–1215.
- Klein, A. G., & Muthén, B. O. (2007). Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*(4), 647–673.
- Kuhn, E., & Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, *49*, 1020–1038.
- Kulikov, G., & Kulikova, M. (2014). Accurate numerical implementation of the continuous-discrete extended Kalman filter. *IEEE Transactions on Automatic Control*, *59*(1), 273–279. doi:10.1109/TAC.2013.2272136.
- Lee, S., & Song, X. (2003). Maximum likelihood estimation and model comparison for mixtures of structural equation models with ignorable missing data. *Journal of Classification*, *20*(2), 221–255. doi:10.1007/s00357-003-0013-5.

- Li, F., Duncan, T. E., & Acock, A. (2000). Modeling interaction effects in latent growth curve models. *Structural Equation Modeling*, 7(4), 497–533.
- Liang, H., Miao, H., & Wu, H. (2010). Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model. *Annals of Applied Statistics*, 4(1), 460–483.
- Longstaff, M. G., & Heath, R. A. (1999). A nonlinear analysis of the temporal characteristics of handwriting. *Human Movement Science*, 18, 485–524.
- Losardo, D. (2012). *An examination of initial condition specification in the structural equation modeling framework*. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill, NC.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44, 190–200.
- Marsh, W. H., Wen, Z. L., & Hau, J.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9, 275–300.
- Mbalawata, I. S., Särkkä, S., & Haario, H. (2013). Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and non-linear Kalman filtering. *Computational Statistics*, 28(3), 1195–1223.
- McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analysis with incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.
- McArdle, J. J., & Hamagami, F. (2003). Structural equation models for evaluating dynamic concepts within longitudinal twin analyses. *Behavior Genetics*, 33(2), 137–159. doi:10.1023/A:1022553901851.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- Miao, H., Xin, X., Perelson, A. S., & Wu, H. (2011). On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*, 53(1), 3–39.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology—this time forever. *Measurement: Interdisciplinary Research and Perspectives*, 2, 201–218.
- Molenaar, P. C. M., & Newell, K. M. (2003). Direct fit of a theoretical model of phase transition in oscillatory finger motions. *British Journal of Mathematical and Statistical Psychology*, 56, 199–214. doi:10.1348/000711003770480002.
- Ortega, J. (1990). *Numerical analysis: A second course*. Philadelphia, PA: Society for Industrial and Academic Press.
- Oud, J. H. L. (2007). Comparison of four procedures to estimate the damped linear differential oscillator for panel data. In J. Oud & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, 65(2), 199–215.
- Oud, J. H. L., & Singer, H. (Eds.). (2010). *Special issue: Continuous time modeling of panel data*, 62 (1).
- Pickering, T. G., Shimbo, D., & Haas, D. (2006). Ambulatory blood-pressure monitoring. *The New England Journal of Medicine*, 354, 2368–2374.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical recipes in C*. Cambridge: Cambridge University Press.
- R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved April, 2014, from <http://www.R-project.org> (ISBN: 3-900051-07-0).
- Ralston, A., & Rabinowitz, P. (2001). *A first course in numerical analysis* (2nd ed.). Mineola, NY: Dover.
- Ramsay, J. O., Hooker, G., Campbell, D., & Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *Journal of Royal Statistical Society: Series B*, 69(5), 741–796.
- Raudenbush, S. W., & Liu, X.-F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6(4), 387–401.
- Särkkä, S. (2013). *Bayesian filtering and smoothing*. Hillsdale, NJ: Cambridge University Press.
- SAS Institute Inc. (2008). SAS 9.2 Help and Documentation (Computer software manual). Cary, NC: SAS Institute Inc.
- Sherwood, A., Steffen, P., Blumenthal, J., Kuhn, C., & Hinderliter, A. L. (2002). Nighttime blood pressure dipping: The role of the sympathetic nervous system. *American Journal of Hypertension*, 15, 111–118.
- Sherwood, A., Thurston, R., Steffen, P., Blumenthal, J. A., Waugh, R. A., & Hinderliter, A. L. (2001). Blunted nighttime blood pressure dipping in postmenopausal women. *American Journal of Hypertension*, 14, 749–754.
- Singer, H. (1992). The aliasing-phenomenon in visual terms. *Journal of Mathematical Sociology*, 14(1), 39–49.
- Singer, H. (1995). Analytical score function for irregularly sampled continuous time stochastic processes with control variables and missing values. *Econometric Theory*, 11, 721–735. doi:10.1017/S0266466600009701.
- Singer, H. (2002). Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood vs. extended Kaman filter and itô-Taylor expansion. *Journal of Computational and Graphical Statistics*, 11, 972–995.
- Singer, H. (2007). Stochastic differential equation models with sampled data. In K. van Montfort, J. H. L. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 73–106). Mahwah, NJ: Lawrence Erlbaum Associates.
- Singer, H. (2010). Sem modeling with singular moment matrices. Part I: MI-estimation of time series. *The Journal of Mathematical Sociology*, 34(4), 301–320. doi:10.1080/0022250X.2010.532259.
- Singer, H. (2012). Sem modeling with singular moment matrices. Part II: MI-estimation of sampled stochastic differential equations. *The Journal of Mathematical Sociology*, 36(1), 22–43. doi:10.1080/0022250X.2010.532259.
- Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (ema) in behavioral medicine. *Annals of Behavioral Medicine*, 16(3), 199–202.

- Strogatz, S. H. (1994). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Cambridge, MA: Westview.
- Stuart, A. M., Voss, J., & Wilberg, P. (2004). Conditional path sampling of sdes and the langevin mcmc method. *Communications in Mathematical Sciences*, 2(4), 685–697.
- Tanizaki, H. (1996). *Nonlinear filters: Estimation and applications* (2nd ed.). Berlin: Springer.
- Thatcher, R. W. (1998). A predator–prey model of human cerebral development. In K. M. Newell & P. C. M. Molenaar (Eds.), *Applications of nonlinear dynamics to developmental process modeling* (pp. 87–128). Mahwah, NJ: Lawrence Erlbaum.
- Wen, Z., Marsh, H. W., & Hau, K.-T. (2002). Interaction effects in growth modeling: A full model. *Structural Equation Modeling*, 9(1), 20–39.
- Wu, H. (2005). Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research*, 14, 171–192.
- Zhu, H., Gu, M., & Peterson, B. (2007). Maximum likelihood from spatial random effects models via the stochastic approximation expectation maximization algorithm. *Statistics and Computing Archive*, 17(2), 163–177.
- Zhu, H. T., & Zhang, H. P. (2006). Generalized score test of homogeneity for mixed effects models. *Annals of Statistics*, 34, 1545–1569.

Manuscript Received: 13 DEC 2013

Published Online Date: 22 NOV 2014