



Published in final edited form as:

Multivariate Behav Res. 2016 ; 51(2-3): 154–184. doi:10.1080/00273171.2015.1123138.

A Comparison of Two-Stage Approaches for Fitting Nonlinear Ordinary Differential Equation (ODE) Models with Mixed Effects

Sy-Miin Chow, Jason J. Bendezú, Pamela M. Cole, and Nilam Ram

The Pennsylvania State University

Abstract

Several approaches currently exist for estimating the derivatives of observed data for model exploration purposes, including functional data analysis (FDA), generalized local linear approximation (GLLA), and generalized orthogonal local derivative approximation (GOLD). These derivative estimation procedures can be used in a two-stage process to fit mixed effects ordinary differential equation (ODE) models. While the performance and utility of these routines for estimating linear ODEs have been established, they have not yet been evaluated in the context of nonlinear ODEs with mixed effects. We compared properties of the GLLA and GOLD to an FDA-based two-stage approach denoted herein as *functional ordinary differential equation with mixed effects* (FODEmixed) in a Monte Carlo study using a nonlinear coupled oscillators model with mixed effects. Simulation results showed that overall, the FODEmixed outperformed both the GLLA and GOLD across all the embedding dimensions considered, but a novel use of a fourth-order GLLA approach combined with very high embedding dimensions yielded estimation results that almost paralleled those from the FODEmixed. We discuss the strengths and limitations of each approach and demonstrate how output from each stage of FODEmixed may be used to inform empirical modeling of young children's self-regulation.

Keywords

Differential equation; functional data analysis; dynamic; dynamical systems; derivatives

Longitudinal data collection and analysis in the social and behavioral sciences have evidenced tremendous growth as researchers expand earlier tools for studying change (e.g., using difference scores; Cronbach & Furby, 1970; Harris, 1963) to better answer – on multiple fronts – whether and in what ways individuals and their behaviors change over time. Across fields, researchers in the physical sciences, econometricians, engineers, and statisticians have all made good use of differential equation models to understand many change phenomena, including the interplay among business firms, species, chemical substrates, and social actors. Along the way, they have developed a variety of analytical methods for fitting differential equation model to single- or multiple-subject time series data (Beskos, Papaspiliopoulos, & Roberts, 2009; Jones, 1993; Mbalawata, Särkkä, & Haario, 2013; Ramsay, Hooker, Campbell, & Cao, 2007; Särkkä, 2013). A growing number of social

scientists, including psychologists and sociologists (e.g., Arminger, 1986; Coleman, 1968; Tuma & Hannan, 1984), have also turned to differential equations to capture the dynamic aspect of social processes, particularly the interdependence among a system of variables.

Much of the work utilizing differential equation models in the psychometric literature has focused on *linear* ordinary and stochastic differential equation (ODE and SDE, respectively) models using exact discrete time and state-space approaches (Oud & Jansen, 2000; Oud & Singer, 2008; Singer, 2010, 2012; Voelkle, Oud, Davidov, & Schmidt, 2012), differential structural equation models (Boker, Neale, & Rausch, 2008), Bayesian approaches (Oravecz, Tuerlinckx, & Vandekerckhove, 2011), and two-stage derivative estimation/modeling approaches (Boker, Deboeck, Edler, & Keel, 2010; Boker & Graham, 1998; Deboeck, 2010; Trail et al., 2013). More difficult, but also gaining traction, are methods for fitting *nonlinear* ODEs and SDEs (e.g., Chow, Lu, Sherwood, & Zhu, 2016, 1; Chow, Ferrer, & Nesselroade, 2007; Lu, Chow, Sherwood, & Zhu, 2015, 3; Molenaar & Newell, 2003; Singer, 2002, 2003). In the present article, we are restricting ourselves to two-stage approaches that facilitate the building and refinement of linear and nonlinear ODEs. Even in these special cases, all of the current routines for handling nonlinear ODEs generally require a level of technical/programming expertise that hampers uptake in applied research settings.¹

The purpose of this paper is to facilitate the study of psychological/behavioral processes through the presentation and demonstration of an accessible two-stage approach for fitting linear and nonlinear differential equation models to empirical data. Our hope is that the approach proposed here, *functional ordinary differential equation with mixed effects* (FODEmixed), might propel discovery and understanding of human behavior by facilitating fitting (and rejection) of many possible models of change quickly and easily. The proposed approach extends prior work and provides new insights in three ways. First, the present article features the first Monte Carlo evaluation of the feasibility of using existing two-stage approaches designed for handling *linear* ODEs – including the generalized local linear approximation (GLLA; Boker et al., 2010) and the generalized orthogonal local derivative (GOLD; Deboeck, 2010) approaches – for fitting *nonlinear* ODEs. Second, we extend prior two-stage, functional data analysis (FDA)-based approaches for derivative estimation using regression splines (Trail et al., 2013) and a nonlinear regression framework (Liang & Wu, 2008) to utilizing penalized regression splines within a nonlinear mixed effects framework. Third, we propose and evaluate the novel use of a 4th-order GLLA with high embedding dimensions against the other estimation approaches considered (i.e., conventional 2nd-order GLLA, GOLD and FODEmixed).

In the sections that follow, we first provide some introductory remarks on the nature and characteristics of ODEs in general, and the specific modeling framework considered in the present article. To motivate the proposed approach, we present several benchmark ODE models and an empirically constructed ODE model of self-regulation subsumed under this

¹Even though some options do exist in standard software packages for fitting nonlinear ODEs, all of these options are plagued by other practical limitations. For instance, the nlmeODE package in R (Tornøe, Agersø, Jonsson, Madsen, & Nielsen, 2004), while allowing users to fit selected ODE models with mixed effects, does not provide users with the option to handle unknown initial conditions; the PROC MODEL routine in SAS (Ferrer & Helm, 2013; SAS Institute Inc., 2008) provides more flexibility in the specifications of the ODEs but does not allow for random effects in the ODE parameters.

framework. We demonstrate, in the context of these examples, the possible utility of two-stage approaches in providing exploratory information for model development and testing purposes. This is followed by an introduction of the two-stage approaches considered in the present article (FODEmixed, GLLA and GOLD) and a simulation study conducted to assess their relative strengths and limitations. We then illustrate the proposed approach with a set of empirical data. We end with a discussion of the key insights from this paper as well as unresolved issues and future directions.

Ordinary Differential Equation (ODE) Models

Formally, differential equation models are representations of systematic relations among derivatives – quantifications of change (and changes in the changes) that are manifested at infinitely small time scales. As a class, differential equation models provide a framework for describing *how* (as opposed to *whether*) people are likely to change. In practice, differential equation models are particularly useful for the study of continuous processes that are observed at regular intervals (e.g., panel and observational studies), or intermittently (e.g., experience sampling, ecological momentary, event-contingent, and other related designs; Hawley, Burlison, Berntson, & Cacioppo, 2003; Merrilees, Goeke-Morey, & Cummings, 2008).

The core building blocks of differential equation models are derivatives. In brief, consider the repeated observations of a continuous process, $x(t)$, obtained over time at times $t = 0, \dots, T$. At any given point in time, the first derivative (rate of change of x) is the change in x that occurs within a (infinitely) small window of time, a . Formally,

$$\frac{dx(t)}{dt} = \lim_{a \rightarrow 0} \frac{x(t+a) - x(t)}{a}.$$

In turn, the second derivative is the change in the rate of change of x . Formally,

$$\frac{d^2x(t)}{dt^2} = \lim_{a \rightarrow 0} \frac{\left[\frac{x(t+a) - x(t)}{a} \right] - \left[\frac{x(t) - x(t-a)}{a} \right]}{a} = \lim_{a \rightarrow 0} \frac{x(t+a) - 2x(t) + x(t-a)}{a^2}.$$

Other higher-order derivatives then capture further changes in these change qualities.

Dynamical phenomena can often be described using ordinary differential equations (ODEs) that indicate how the derivatives and levels of different variables are related to each other. The ODEs may be expressed in a vector form as:

$$\frac{dx_i(t)}{dt} = \mathbf{f}[x_i(t), \boldsymbol{\theta}_{f,i}, t], \quad i=1, \dots, n; t=0, \dots, T; \quad (1)$$

where i indexes person and t indexes time, $\mathbf{f}[\cdot]$ is a vector of drift functions; $\boldsymbol{\theta}_{f,i}$ represents a $q \times 1$ vector of person-specific parameters that affect the dynamic functions in Equation 1;

$\mathbf{x}_i(t)$ is an $n_x \times 1$ vector of latent variables of interest containing level and derivatives at time t that are of lower order than those in $\frac{d\mathbf{x}_i(t)}{dt}$; $\frac{d\mathbf{x}_i(t)}{dt}$ is a corresponding $n_x \times 1$ vector of derivatives one order higher than those in $\mathbf{x}_i(t)$ (e.g., level as the 0-order derivative is one order lower than the first derivative; the first derivative is one order lower than the second derivative, etc.). We further partition $\mathbf{x}_i(t)$ into two subvectors as $\mathbf{x}_i(t) = [\mathbf{x}_i^{(1)}(t)' \ \mathbf{x}_i^{(2)}(t)']'$, where $\mathbf{x}_i^{(1)}(t)$ is an $n_y \times 1$ subvector containing all the level variables in the ODE model and $\mathbf{x}_i^{(2)}(t)$ is an $n_x - n_y$ subvector containing all the lower-order derivative variables. Thus, for the h th modeled process of order m_h , $\mathbf{x}_i^{(2)}(t) = [dx_{h,i}^{(1)}(t) \dots d^{m_h-1}x_{h,i}^{(1)}(t)]'$. The order of an ODE, denoted as m below, then indicates the highest-order derivative that appears in the model (i.e., the maximum of all values of m_h). If all functions in $f[\cdot]$ are linear – namely, the relations among the dependent variables and their derivatives are additive and each term is raised to no higher than the first power – then Equation 1 is said to be a linear ODE model. Note that even though linear ODEs typically have nonlinear integral solutions (i.e., produce nonlinear trajectories), nonlinearities refer specifically to the presence of nonlinear terms in the ODE functions in Equation 1.

Individual differences in $\theta_{f,i}$ are modeled in the present context through a linear mixed effects model,

$$\theta_{f,i} = \mathbf{H}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i; \quad (2)$$

where $\boldsymbol{\beta}$ is a $p_\beta \times 1$ vector of fixed effects parameters, and \mathbf{b}_i a $d \times 1$ vector of random effects; \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. Typically, \mathbf{b}_i is assumed to be multivariate normally distributed, $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$, with θ_b containing all the unknown parameters in $\boldsymbol{\Sigma}_b$.

In the present context, we restrict ourselves to a relatively simple measurement process: one in which a single manifest indicator is used to identify each underlying ODE process of interest. Specifically, we assume that the level variables in $\mathbf{x}_i^{(1)}(t)$ are all latent true scores that are each indicated by only one manifest variable at individual-specific but possibly irregularly spaced time points, $t = t_{i,j}, j = 1, \dots, T_i$, with $\Delta_{i,j} = t_{i,j+1} - t_{i,j}$. At each observed time point, there is an $n_y \times 1$ vector that is linked to $\mathbf{x}_i^{(1)}(t_{i,j})$:

$$\mathbf{y}_i(t_{i,j}) = \mathbf{x}_i^{(1)}(t_{i,j}) + \boldsymbol{\varepsilon}_i(t_{i,j}); \quad (3)$$

where $\boldsymbol{\varepsilon}_i(t_{i,j})$ denotes multivariate normally distributed measurement error processes such that $E[\boldsymbol{\varepsilon}_i(t_{i,j})] = \mathbf{0}$, $E[\boldsymbol{\varepsilon}_i(t_{i,j})\boldsymbol{\varepsilon}_i(t_{i,j})'] = \boldsymbol{\Sigma}_\varepsilon$, $\boldsymbol{\varepsilon}_i(t_{i,j})$ and $\mathbf{x}_i(t_{i,j})$ are independent, and $\boldsymbol{\varepsilon}_i(t_{i,j})$ and $\boldsymbol{\varepsilon}_i(t_{i,k})$ are independent for $t_{i,j} \neq t_{i,k}$, with a diagonal structure for $\text{Cov}(\boldsymbol{\varepsilon}_i(t_{i,j})) = \boldsymbol{\Sigma}_\varepsilon$. Furthermore, $\mathbf{y}_i = \{\mathbf{y}_i(t_{i,j}), j = 1, \dots, T_i\}$ is an n_y -variate time series for person i , and θ_ε is a $p_\varepsilon \times 1$ vector containing all the unknown parameters in $\boldsymbol{\Sigma}_\varepsilon$. Throughout, we use the notation

(t) to denote any time point in general whereas t_{ij} is used to denote the j th observed time point for person i .

Examples of ODEs

ODEs are used to describe many types of change phenomena. In this section we review several benchmark examples of ODE models that are either well-known in the dynamical systems literature, or have close correspondence to other widely adopted models in the psychometric literature. Using these models, we demonstrate some possible ways in which exploratory information from two-stage procedures may be utilized to aid model developments and adaptations.

One of the simplest ODE models is one in which the rate of change of a process is held constant over time as:

$$\frac{dx_i(t)}{dt} = a_{1i}, \quad (4)$$

where a_{1i} is a person-specific scalar value that describes the constant growth/decline rate for individual i . Integrating the ODE in (4) we obtain a solution as

$$x_i(t) = x_{0i} + a_{1i}t, \quad (5)$$

where x_{0i} is a coefficient of integration that represents the *initial condition* for person i 's dynamic process at time $t = 0$. As discussed by Ram and Grimm (2015), the ODE in Equation 4, with the addition of measurement errors, is the well-known linear growth curve model (Meredith & Tisak, 1990). Of note, even though the person-specific initial condition, x_{0i} , here coincides with the random intercept in the linear growth curve model, the initial conditions for an ODE generally include the initial values for all the variables in $\mathbf{x}(t)$, including both levels (corresponding to “intercepts”) and other derivative variables. Many ODE applications assume that all the initial conditions are known a priori or have to be specified explicitly. One merit of all the two-stage procedures is that they preclude the need to estimate or specify the initial conditions of an ODE, thus easing estimation difficulties.

A slightly more complex ODE model that is often cited in the sociological literature (e.g. Coleman, 1968; Nielson, 1980; Rosenfeld, 1980) is a proportional change model,

$$\frac{dx(t)}{dt} = c(x^* - x(t)), \quad (6)$$

where the first derivative is a linear function of $x(t)$, the current level of x , and x^* , an asymptote of the system; c is a growth/decline rate parameter. Integration with respect to t yields the solution expressed as

$$x(t) = x_0 \exp(ct) + x^*(1 - \exp(ct)), \quad (7)$$

where x_0 is again a coefficient of integration that represents the initial condition at $t = 0$. When c has a negative value, the exponential terms approach zero and $x(t)$ approaches the asymptotic equilibrium given by x^* .² The system in Equation 6 has two equilibrium levels of $x(t)$, at which point $\frac{dx(t)}{dt} = 0$. The equilibrium points occur at $x(t) = x^*$, or when c is equal to zero. For illustration, plots of the solution trajectories over time, $x(t)$, generated using a numerical solver with Equation (6) with $x^* = 3$ and three values of c (0, -0.3, -1), are shown in Figure 1(A). We also plotted the *phase plane* of this ODE – a plot of some of the variables included in the ODE system, in this case $\frac{dx(t)}{dt}$ and $x(t)$. The phase plane plot highlights that the system is static (i.e., intersects with the dot-dashed horizontal line, $\frac{dx(t)}{dt} = 0$) at $x(t) = x^*$ and when $c = 0$, in which case there is only a single dot, and the first derivative stays at zero throughout.

While ODEs such as those in Equations 4 and 6 may be used to describe a wide variety of growth and diffusion processes (Banks, 1994), other ODEs describe oscillatory processes. For example, a damped linear oscillator model (Boker & Graham, 1998) can be used to describe behavior that ebbs and flows like the swinging of a pendulum with friction,

$$\frac{d^2 x_1(t)}{dt^2} = \eta x_1(t) + \zeta \frac{dx_1(t)}{dt}, \quad (8)$$

where $x_1(t)$ represents the level or displacement of the pendulum relative to its equilibrium position (the center of motion); η is a frequency parameter that governs how rapidly the pendulum swings back and forth relative to the equilibrium point and ζ is a parameter that controls the extent to which the pendulum shows damping (if $\zeta < 0$) or amplified motion (if $\zeta > 0$). Like the models above, Equation 8 has analytic solutions; the mathematical form of the solution depends on the values of η and ζ (Hu, Boker, Neale, & Klump, 2014), but all solutions can be readily found in most standard textbooks on ODEs (e.g., Zill, 1993), or in discrete-time form for the SDE variation of Equation 8 (Oud, 2007; Voelkle et al., 2012).³

Plots of $x(t)$ over time generated from Equation (8) with $\eta = -0.8$ and three values of ζ (-0.01, -0.1, -1.5) are shown in Figure 1(C) along with the phase plane plots in panels (D) and (E) using the same initial conditions ($x_1(0) = 2$ and $\frac{dx_1(0)}{dt} = -2$ for all three trajectories.).

²Growth curve modelers may recognize this system as the constrained exponential growth model (see Ram & Grimm, 2015).

³Any higher-order ODEs (i.e., ODEs that involve derivatives of order beyond 1) can be rewritten into a system of multiple first-order ODEs. For instance, the second-order ODE in Equation (8) can be re-expressed in vector form as depicted in Equation 1, thereby featuring a vector of two first-order ODE functions as

$$\frac{d}{dt} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ \eta x_1(t) + \zeta x_2(t) \end{bmatrix}. \quad (9)$$

While the linear relations between levels and second derivatives are evident in Figure 1(D), the magnitudes of the slopes (corresponding to η) appear to differ from -0.8 . The correct relation would be observed, however, if the effects of $\frac{dx_1(t)}{dt}$ are controlled for. To this end, standard regression diagnostic tools can be used to detect partial associations among variables after controlling for the effects of other predictors. One such tool is the component-plus-residuals plot (Faraway, 2004; Fox, 2009) – a plot of the residuals of the dependent variable against a predictor after the effects of other predictors have been partialled out; a Loess line and a linear least squares line are overlaid on the plot to help visualize any possible deviations in associations from linearity (Faraway, 2004; Fox, 2009). Here, the component-plus-residual plot in panel (F) reveals that a negative linear association is likely sufficient for capturing the relation between $d^2x(t)/dt^2$ and $x(t)$, and that this association did not deviate notably for different cases in the data set. In contrast, the partial relation between $d^2x(t)/dt^2$ and $dx(t)/dt$ depicted in panel (G) highlights the possibility that there are three different relations embedded in the data: two negative slopes of different degrees of steepness, and another set of points that is basically flat. In sum, these plots and other related diagnostic tools (see e.g., Faraway, 2004; Fox, 2009; Harrell, 2001) help identify the true model that generated the data – useful tools when the true model is unknown.

Intensive longitudinal data provide opportunities to examine a wide variety of dynamic processes that drive many aspects of individual behavior and development. Many of these processes (e.g., self-regulation, dyadic interaction) are filled with state-dependent changes wherein changes in one process are dependent on the states, or unobserved latent values of other processes. For instance, nonlinear ODEs have been used to model, among other phenomena, ovulatory regulation (Boker, Neale, & Klump, 2014), circadian rhythms (Brown & Luithardt, 1999), cerebral development (Thatcher, 1998), substance use (Boker & Graham, 1998), cognitive aging (Chow & Nesselroade, 2004), parent-child interactions (Thomas & Martin, 1976), dyadic relationships (Chow et al., 2007); and sudden transitions in attitudes (van der Maas, Kolstein, & van der Pligt, 2003).

One well-known nonlinear ODE model is the Lotka-Volterra model (Lotka, 1925; Volterra, 1926), often used to describe the dynamics of predator and prey populations as:

$$\begin{aligned}\frac{dx_1(t)}{dt} &= r_1x_1(t) - a_{12}x_1(t)x_2(t), \\ \frac{dx_2(t)}{dt} &= -r_2x_2(t) + a_{21}x_1(t)x_2(t),\end{aligned}\quad (10)$$

where $x_1(t)$ and $x_2(t)$ are the densities of the prey and predatory populations, respectively, at time t , r_1 is the growth rate of the prey population, and r_2 is the death rate of the predator population. Interactions between the predator and prey reduce the density of the prey population and increase in the density of the predator population as noted the signs of a_{12} and a_{21} . With a particular range of parameters (and non-zero initial conditions) the predator-prey model produces lead-lag cyclic fluctuations in the predator and prey densities like those shown in Figure 2(A) and (B). Adaptations of this model have been used to represent a variety of behavioral and psychological phenomena. For example, Chow and Nesselroade

(2004) developed a model for age differences in susceptibility to cognitive interference by including negative quadratic components, $-a_{11}x_1^2(t)$ and $-a_{22}x_2^2(t)$ that accommodate additional “intra-species” competition that manifests prominently when population density is high. As shown in Figure 2(C) and (D), including a negative quadratic component, $-0.20x_1^2(t)$, to the prey equation in Equation 10, creates some intrinsic damping of the population peaks. Other adaptations of the classic predator-prey model include the “piecewise” linear versions used by Gottman and colleagues to represent marital behavior (Gottman, Murray, Swanson, Tyson, & Swanson, 2002), in which the “inter-species”/inter-spousal couplings are replaced by simple linear functions (e.g. $a_{12}x_2(t)$ and $a_{21}x_1(t)$) or other theoretically interesting influence functions (Madhyastha, Hamaker, & Gottman, 2011; Ram, Shiyko, Lunkenheimer, Doerksen, & Conroy, 2014). The main point being that models established in other areas can be adapted and/or reconfigured in ways that accommodate the specific theories and data of interest.

One final nonlinear ODE example is a model first introduced by B. van der Pol and J. van der Mark to describe cardiac oscillations (cf., Kaplan & Glass, 1995), and extended further to model human circadian rhythms (Brown & Luithardt, 1999; Chow et al., 2016, 1). In basic form the van der Pol oscillator model is written as

$$\frac{d^2x(t)}{dt^2} = \eta x(t) + \zeta [1 - x^2(t)] \frac{dx(t)}{dt}. \quad (11)$$

Similar in construction to the linear damped oscillator model in Equation 8, the van der Pol oscillator also models the second derivatives, $\frac{d^2x(t)}{dt^2}$, as a function of the current level, $x(t)$, and current rate of change, $\frac{dx(t)}{dt}$. As in the linear model, η governs the frequency of the oscillatory process (and often set to -1), and ζ is a damping or amplification parameter that governs the amplitude of the oscillations. However, here the inclusion of $x^2(t)$ as a multiplier of $\frac{dx(t)}{dt}$ means that the extent of damping (or amplification) is moderated by the squared level of $x(t)$. Thus, when $\zeta_j > 0$, wherein the amplification at small values of $x(t)$ turns into large damping at extreme values of $x(t)$. The result is a system characterized by saw-tooth type oscillations (Strogatz, 1994), shown in Figure 2(E) for systems with $\eta = 1$ and three values of ζ (1.0, 3.0, 5.0).

The phase plane plot of the van der Pol oscillator model is shown in Figure 2(F). The plot clearly shows that the relation between the second and first derivatives is nonlinear. If we did not know the underlying model, this would be a first indication that nonlinear terms may need to be included. The component + residuals plot in panel (G) shows the relation between the second and first derivatives after the linear effect of $x(t)$ has been partialled out for the case where $\zeta = 1.0$. The slope of the linear regression of $dx(t)/dt$ on $d^2x(t)/dt^2$ is near zero. However, the Loess regression line indicates the presence of some nonlinearities. Furthermore, the “bow-tie-like” like patterns of association between $dx(t)/dt$ on $d^2x(t)/dt^2$ suggest the possibility that this relation is moderated by another process that is large while away from the origin (a point where the level, first and second derivatives are all equal to

zero) but small near the origin. One possible model is to allow the effect of $dx(t)/dt$ on $d^2x(t)/dt^2$ be moderated by $x^2(t)$ because the magnitudes of $x^2(t)$ are, by definition, small when $x(t)$ is close to zero and large when $x(t)$ is far from zero. Indeed, the correlation between the product term, $x^2(t)(dx(t)/dt)$, and the residuals from a model where only the linear effects of $x(t)$ and $dx(t)/dt$ were included, was strong ($r = -0.66$), whereas the correlation between $x(t)(dx(t)/dt)$ and those residuals was weak (e.g., $r = -0.09$). After incorporating the nonlinear term $x^2(t)dx(t)/dt$ into the baseline linear model (see panel (H)), the Loess smoother line is now near zero, and while the residuals still have some discernable patterns of nonlinearities, much of the nonlinear relation seen in panel (G) had now been accounted for.

In sum, the move into the nonlinear modeling realm opens up a myriad of new opportunities to represent within-person dynamics and interindividual differences therein. However, the increase in possibility also makes examining, sorting, and iterating through a rich array of potentially useful models a daunting task. We seek to underscore three additional points here. First, even in cases involving linear ODEs, options for pursuing model explorations and estimation may not be known to many social and behavioral scientists.⁴ Nonlinear ODEs of the form in in Equations 1–3 generally do not have analytic solutions and are associated with added computational difficulties. Most of the existing single-stage approaches for estimating nonlinear ODEs/SDEs with random effects (e.g., Chow et al., 2016, 1; Singer, 2012) require appropriate handling of the initial conditions of the latent variables; possible use of numerical integration techniques to obtain approximation solutions to the ODEs, as well as the use of computational techniques (e.g., Monte Carlo methods) to perform integration over the random effects in \mathbf{b}_j (Beskos et al., 2009; Mbalawata et al., 2013; Särkkä, 2013). The two-stage estimation procedures considered in the present article use FDA methods, the GLLA, or the GOLD to obtain approximations of all level and derivative terms, so that Equations 1–3 reduce to a linear or nonlinear mixed effects model. The implementation process is thus greatly facilitated by estimation procedures already available in many standard software packages.

Second, graphical summaries of the phase space of a system can help researchers conceptualize, build, and identify a viable model. When the underlying model is known, derivative values can be computed directly from the ODEs. Otherwise, some kind of numerical differencing/differentiation procedures (like the FDA, GLLA and GOLD approached examined in this paper) might be used to obtain the derivative estimates. Third, parallel to the confounds that may arise from omitting linear terms from a regression model with nonlinear terms (Faraway, 2004), the same practical rules may be used in the pursuance of ODE models such that linear terms are included prior to incorporating nonlinear ones. However, even with that, it is not always clear what forms of nonlinearities a researcher should consider and sample size constraints may limit the number of predictors a researcher can realistically consider. We have, for illustration, greatly simplified the actual process of

⁴For instance, in earlier approaches of fitting the proportional change model, several researchers fitted reparameterized versions of Equation (7) to the data without imposing the appropriate nonlinear constraints, a practice that could lead to severely biased parameter estimates (Hamerle, Nagl, & Singer, 1990). Other approaches that do impose the appropriate nonlinear constraints, such as those based on the exact discrete time approach, have been proposed since then (Hamerle et al., 1990; Oud & Jansen, 2000). Unfortunately, uptake in applied settings remains slow.

searching for appropriate nonlinear terms. In practice, one must carefully consider the influence of each term – linear or nonlinear – and build the model through consideration of both theoretical and statistical fit.

Motivating Example

One core assumption of most models of self-regulation is that it entails the influence of *executive processes* – namely, self-initiated regulatory attempts – on *prepotent responses* (i.e., frustration, desire, immediate gratification). Despite its being conceptualized as a dynamic process that inherently involves change (Boker & Laurenceau, 2006; Cole, Martin, & Dennis, 2004), most early childhood self-regulation studies focus on between-child comparisons utilizing summary measures aggregated across time (e.g., latency and frequency counts of desired or undesired behaviors). These methods obscure temporal dynamics which, when placed within the appropriate analytic framework, might reveal subtle nuances in the overall effectiveness of children’s self-regulation attempts. In order for self-regulation to be effective, executive process engagement has to override children’s natural or prepotent responses to satisfy their immediate desires over time. In parallel, it is also possible for children’s pre-potent responses to show reductions or increases in amplitude (damping and amplification, respectively) over the course of a demanding situation for reasons that are unrelated to executive efforts (e.g., fatigue). Of particular interest to developmental psychologists is whether and how such changes in prepotent response amplitudes depend on – or interact with – the child’s current level of executive efforts to damp or amplify their prepotent responses. Nonlinear ODE models can explicitly capture this dynamic tension as nonlinear interactive processes that unfold over time. However, such modeling endeavors have been challenging. One difficulty is that despite the call for dynamic approaches to studying child self-regulation (Cole et al., 2004), there are no studies that avail themselves of the full time course of the observational tasks that are used. Another challenge is that developmental researchers are still trying to identify appropriate candidate models for demonstrating these dynamic processes, and tools in the psychometric literature for fitting nonlinear ODE models are scarce. In the empirical example section, we demonstrate how the FODEmixed approach facilitates such model exploration and model fitting in an investigation of young children’s self-regulation dynamics.

The Functional Ordinary Differential Equation with Mixed Effects (FODEmixed) Approach

We introduce FODEmixed, as a two-stage approach for estimating linear and non-linear differential equation model. In Stage 1 FDA methods – a repertoire of tools for flexibly approximating and analyzing curves and their corresponding derivatives (Ramsay & Silverman, 2005) – are used to obtain smoothed level and derivative estimates. Implementation is facilitate by the “fda” package in R (Ramsay, Hooker, & Graves, 2009). In Stage 2, smoothed estimates from Stage 1 are used as observed variables in a linear or nonlinear mixed effects model (e.g., “nlme” package in R). Details for each of the two stages follow.

Stage 1: Derivative estimation using FDA

Given a time series for the h th manifest variable, one possible way to approximate the ebbs and flows in the time series is to use a collection of basis splines (B-splines; De Boor, 1977, 1978; Dierckx, 1993) composed of K functions as:

$$\hat{y}_{h,i}(t) = \hat{x}_{h,i}^{(1)}(t) = \sum_{k=0}^K c_k \phi_{k,i}(t), \quad (12)$$

where $\phi_{k,i}(t)$ is the k th known basis function of time for person i , and c_k is its associated weight or coefficient. More generally, B-splines can be understood as a way of implementing piecewise polynomials up to degree K within segments of a time series (De Boor, 1977, 1978; Dierckx, 1993). Each segment of the time series is separated from an immediately adjacent segment by a *knot point*, with the first and last measurement occasions typically constituting the outer or ending knots. To ensure that a particular approximation curve is connected smoothly at an interior knot point, two adjacent polynomials are typically required to match in the values of a fixed number of their derivatives, usually chosen to be $K - 1$ (see e.g., Cudeck & Klebe, 2002).⁵ The number of basis functions used in each segment, also known as the order of a B-spline, is equal to $K + 1$. It is customary to specify the order of the B-spline to be at least two higher than the order of the derivative estimates of interest or alternatively, the order of the derivatives invoked in the estimation process, whichever one is higher. In this paper, we use a B-spline of order 6 because the estimation of the second-order ODEs considered in this article requires the use of fourth derivatives in the estimation process.

Regularization refers to the process of imposing some kind of roughness penalty on the estimation routines to ensure that the approximated curves satisfy some notion of smoothness (Ramsay & Silverman, 2005). One way of doing so is to estimate the approximation curve (i.e., to estimate the basis function coefficients, c_k) for the h th variable by minimizing the penalized residual sum of squares function

$$PENSSSE_{\lambda,h} = \sum_i [y_{h,i}(t_{i,j}) - \hat{y}_{h,i}(t_{i,j})]^2 + \lambda_h PENALTY(\hat{y}_{h,i}) \quad (13)$$

where $PENALTY(\hat{y}_{h,i})$ is a penalty function that captures the extent of deviations from a predefined smoothness criterion; $\lambda_h \geq 0$ is a smoothing parameter that has to be estimated, or selected using selection criteria such as the generalized cross-validation index (GCV; lower is better) or information criterion measures (Shiyko, Lanza, Tan, Li, & Shiffman, 2012; Tan, Shiyko, Li, Li, & Dierker, 2012). The larger λ_h is, the heavier the penalty (i.e., the estimated curve is smoother). The formulation above assumes that the basis coefficients, c_k ($K = 1, \dots$,

⁵For example, a spline of degree 0 is a constant (a step function) that is discontinuous at knots; a spline of degree 2 is piecewise quadratic with matching level and first derivative (slope) at the interior knot points, and a cubic spline is piecewise cubic with 1st and 2nd derivatives match at knot points to give visibly smooth approximation curves.

K), and smoothing parameter, λ_h , are held invariant across individuals. Alternatively, regularized smoothing may be implemented at the individual level, with person-specific $c_{k,i}$ and $\lambda_{h,i}$ for each participant, to obtain approximation curves, $\hat{y}_{h,i}(t_{i,j})$, that may better capture the patterns of variability in each individual's time-series.

One penalty function typically used for derivative estimation purposes is

$$PENALTY(\hat{y}_{h,i}) = \int \left(\frac{d^{m+2}\hat{y}_{h,i}(s)}{ds^{m+2}} \right)^2 ds, \quad (14)$$

where m is the highest derivative desired. All the ODE models considered in the present article utilize only the level, first and second derivatives. Thus, with $m = 2$, the fourth derivative is used in the penalty function (14) to ensure smoothness of the second derivative. As such we use B-splines of order 6 (two higher than the order of the derivatives invoked in the estimation process) in all our simulation and empirical examples.

Selecting the number and placement of knot points are among some of decisions the users have to make to determine the complexity and smoothness (or alternatively, "wiggleness") of the approximated curves. There are some automated schemes that help guide these decisions (e.g., Eilers & Marx, 1996; Tan et al., 2012; Wood, 2003). However, simultaneous estimation of the smoothing parameter and knot-point related properties is a challenge. Thus, in the simulation and empirical examples considered in the present article, we place the knot points, a priori, at the observed measurement occasions and choose the smoothing parameter, λ_h , by minimizing the GCV function.

After selecting the order of the basis functions, knot points, and an initial value for the smoothing parameter, the basis coefficients of the penalized approximation curves are estimated by minimizing the penalized residual sum of squares function given in Equations 13–14. Estimation of the basis coefficients and smoothing parameter is then repeated iteratively until reasonable approximations of individual curves are obtained. From there, once $\hat{x}_{h,i}^{(1)}(t)$ is available from Equation 12, estimation of derivatives of order p can be obtained as:

$$\frac{d^p \hat{x}_{h,i}^{(1)}(t)}{dt^p} = \frac{d^p \hat{y}_{h,i}(t)}{dt^p} = \sum_{k=0}^K c_k \frac{d^p \phi_{k,i}(t)}{dt^p}, \quad (15)$$

from which we obtain $\hat{\mathbf{x}}_{h,i}^{(2)}(t) = \left[\frac{d\hat{x}_{h,i}^{(1)}(t)}{dt} \dots \frac{d^{m_h-1}\hat{x}_{h,i}^{(1)}(t)}{dt^{m_h-1}} \right]^T$, $\hat{\mathbf{x}}_i^{(2)}(t) = [\hat{\mathbf{x}}_{1,i}^{(2)}(t) \dots \hat{\mathbf{x}}_{n_y,i}^{(2)}(t)]^T$,

and $\frac{d^{(m)} \hat{\mathbf{x}}_i(t)}{dt^{(m)}} = \left[\frac{d^{m_1} \hat{x}_{1,i}^{(1)}(t)}{dt^{m_1}} \dots \frac{d^{m_{n_y}} \hat{x}_{n_y,i}^{(1)}(t)}{dt^{m_{n_y}}} \right]^T$. In particular, $\frac{d^{(m)} \hat{\mathbf{x}}_i(t)}{dt^{(m)}}$ is a vector consisting of approximations of the highest-order derivatives implicated in the ODEs of the modeled

processes, i.e., $\frac{d^m \hat{x}_{h,i}^{(1)}(t)}{dt^{m_h}}$, for $h = 1, \dots, n_y$. These highest-order derivatives at the observed time points, $\frac{d^m \hat{x}_{h,i}^{(1)}(t_{i,j})}{dt^{m_h}}$, $j = 1, \dots, T$, are subsequently used as the “dependent variables” in a mixed effects model in Stage 2.

Once obtained, the smoothed level and derivative estimates can be plotted with respect to time and one another (yielding phase plane plots), and examined for systematic patterns that help guide subsequent selection of linear and/or nonlinear ODE models as shown previously. We note two practical advantages of using FDA-based methods for derivative estimation and model exploration purposes. First, once c_k is known, smoothed level and derivative estimates can be obtained for any given time point because $\varphi_{k,i}(t)$ is a known function of time. As such, this approach can be used with either equally spaced or irregularly spaced time series data. Second, relatively well established guidelines exist to aid decisions on the order of the basis functions, knot points, and smoothness parameter, and these decisions can be made with relatively little dependence on the nature and forms of the true underlying processes.

Stage 2: Fitting mixed effects ODE model

In Stage 2, the smoothed level and derivative estimates from Stage 1, namely,

$\hat{\mathbf{x}}_i^{(1)}(t_{i,j})$, $\hat{\mathbf{x}}_i^{(2)}(t_{i,j})$ and $\frac{d^{(m)} \hat{\mathbf{x}}_i(t)}{dt^{(m)}}$ for all the observed time points (i.e., $j = 1, \dots, T$) are used as observed variables to fit the hypothesized mixed effects ODE of interest as:

$$\begin{aligned} \frac{d\hat{\mathbf{x}}_i^{(m)}(t)}{dt^{(m)}} &= \mathbf{f}[\hat{\mathbf{x}}_i^{(1)}(t_{i,j}), \hat{\mathbf{x}}_i^{(2)}(t_{i,j}), \boldsymbol{\theta}_{f,i}, t_{i,j}] + \boldsymbol{\varepsilon}_i^*(t_{i,j}), \\ \boldsymbol{\theta}_{f,i} &= \mathbf{H}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \quad \boldsymbol{\varepsilon}_i^*(t_{i,j}) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon}^*); \quad \mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b) \end{aligned} \quad (16)$$

where $\boldsymbol{\varepsilon}_i^*(t_{i,j})$ is a vector of residuals from model fitting that may show some variations (e.g., attenuation) in magnitudes compared to the measurement errors in $\boldsymbol{\varepsilon}_f(t_{i,j})$ due to possible changes incurred on the true change trajectories through the smoothing procedures. To utilize standard mixed effects estimation routine implemented in software packages (e.g., the *nlme* function in R; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2014), multivariate ODE processes such as those posited in the classical predator-prey model (Lotka, 1925; Volterra, 1926) or the coupled oscillators model (Boker & Graham, 1998) may be implemented by using the dummy indicator approach as described e.g., in MacCallum, Kim, Malarkey, and Kiecolt-Glaser (1997). Estimation of the parameters in $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_{\varepsilon}^*$ and $\boldsymbol{\Sigma}_b$ can then be obtained using maximum likelihood or restricted maximum likelihood methods as described e.g., in Lindstrom and Bates (1990).

Local Linear Approximation (LLA)-Based Derivative Estimation Approaches

In evaluating the utility of the approach described above, we compare the performance of the proposed FODEmixed approach to two other approaches, namely, Boker and colleagues' (2010) GLLA and Deboeck's (2010) GOLD (proposed as an improvement over the GLLA) for obtaining derivative estimates in Stage 1 (with Stage 2 proceeding as described above). In brief, these two approaches use time delay embedding of the data and a matrix of

predefined weights, \mathbf{W}_h to obtain derivative estimates. In GLLA (but not GOLD), the weights for calculating derivative estimates depend on the maximum order of derivative being estimated. We first describe the time delay embedded data matrix to be used in both the GLLA and GOLD, and then outline the specifics for each approach.

The first step in GLLA and GOLD is to construct a time-delay embedded matrix, $\mathbf{Y}_\tau^{(d)}$, that contains lagged replications of the original time series. The structure of this matrix depends on a user selected d , the embedding dimension (i.e., the number of lags used for derivative estimation), and τ , the spacing between occasions that constitute successive lags. Each individual i 's observed and lagged time series for the h th manifest variable then constitutes a submatrix of $T_i - (d-1)\tau$ rows and d columns, stacked together vertically with other

participants' submatrices to yield $\mathbf{Y}_{h,\tau}^{(d)}$, a d -order time delay embedded matrix with

$\sum_i^n T_i - (d-1)n\tau$ rows and d columns from all participants. For instance, when $d=3$, $\mathbf{Y}_{h,\tau}^{(3)}$ for the h th observed variable, $y_{h,i}(t_{i,j})$ is given by

$$\mathbf{Y}_{h,\tau}^{(3)} = \begin{bmatrix} \mathbf{y}_h(t-\tau) & \mathbf{y}_h(t) & \mathbf{y}_h(t+\tau) \\ y_{h,i}(t_{1,1}) & y_{h,i}(t_{1,1+\tau}) & y_{h,i}(t_{1,1+2\tau}) \\ \vdots & \vdots & \vdots \\ y_{h,i}(t_{1,T_i-2\tau}) & y_{h,i}(t_{1,T_i-\tau}) & y_{h,i}(t_{1,T_i}) \\ \vdots & \vdots & \vdots \\ y_{h,i}(t_{n,1}) & y_{h,i}(t_{n,1+\tau}) & y_{h,i}(t_{n,1+2\tau}) \\ \vdots & \vdots & \vdots \\ y_{h,i}(t_{n,T_n-2\tau}) & y_{h,i}(t_{n,T_n-\tau}) & y_{h,i}(t_{n,T_n}) \end{bmatrix} \quad (17)$$

The full time delay embedded matrix then consists of $\mathbf{Y}_\tau^{(d)} = [\mathbf{Y}_{1,\tau}^{(d)} \dots \mathbf{Y}_{n_y,\tau}^{(d)}]$.

When using GLLA and GOLD users must select τ and d . In earlier implementations of the GLLA, the selection of τ was informed by fitting the hypothesized ODE model multiple times to GLLA-based derivative estimates obtained across a range of τ values, and identifying the optimal τ as the τ that provides the best model fit (e.g., in terms of R^2 , or information criterion measures), or the τ at which the estimate of particular model parameter (e.g., the frequency parameter in a damped linear oscillator model) has begun to stabilize (Hu et al., 2014). More recently, there has been a shift toward setting τ to 1 and selecting d such that the embedding dimension spans further than notable autocorrelations in the time series (Boker et al., 2014). In line with this newer, "model-free" approach, we set τ to 1 and compare the performance of the GLLA and GOLD to the FDA approach over a range of d values.

Stage 1 Using the Generalized Local Linear Approximation (GLLA) Approach

The GLLA approach uses polynomials from Taylor series expansion to obtain estimates of local derivatives for time-series data. Implementation is facilitated by the R functions available for download at <http://people.virginia.edu/smb3u/GLLAfunctions.R>.

The GLLA generalizes its precursor, the local linear approximation (LLA) approach (Boker & Graham, 1998), in which the columns of the time delay embedded data matrix (e.g., $y_h(t - \tau)$, $y_h(t)$, and $y_h(t + \tau)$ for the h th variable) are used to compute the first and second derivatives as

$$\begin{aligned} \frac{dy_h(t)}{dt} &= \frac{y_h(t+\tau) - y_h(t-\tau)}{2\tau} \\ \frac{d^2y_h(t)}{dt^2} &= \frac{y_h(t+\tau) - 2y_h(t) + y_h(t-\tau)}{\tau^2} \end{aligned} \quad (18)$$

Generalizing the LLA, Boker et al. (2010) showed that the differencing weights can be gathered into a matrix, denoted as $\mathbf{W}_{h,GLLA}$, to compute approximate derivatives up to any order m_h by computing $\tilde{D} \hat{\mathbf{Y}}_{h,\tau}^{(d)} = [\hat{y}_h(t) \frac{d\hat{y}_h(t)}{dt} \dots \frac{d^{m_h} \hat{y}_h(t)}{dt^{m_h}}]$ as

$$\tilde{D} \hat{\mathbf{Y}}_{h,\tau}^{(d)} = \mathbf{Y}_{h,\tau}^{(d)} \mathbf{W}_{h,GLLA} = \mathbf{Y}_{h,\tau}^{(d)} \mathbf{L}_h (\mathbf{L}_h' \mathbf{L}_h)^{-1} \quad (19)$$

where \mathbf{L}_h is a $d \times (m_h + 1)$ matrix such that the k th column is given by

$$\mathbf{L}_{h,k} = \frac{[\tau(\nu - \bar{\nu})]^{(k-1)}}{(k-1)!}, \quad \nu = [1, \dots, d]' \quad (20)$$

and ν is a vector of integers from 1 to d . All the columns beyond the first column of $\tilde{D} \hat{\mathbf{Y}}_{h,\tau}^{(d)}$ contain $\frac{d^p \hat{y}_h(t)}{dt^p}$, $p = 1, \dots, m_h$, namely, the least squares estimates of the p th derivative for time series h obtained via multiplication with $\mathbf{W}_{h,GLLA}$, a weight matrix assumed to be known (i.e., not estimated, as distinct from the FDA). The original GLLA set-up was designed for equally spaced data because it assumes that the same set of time intervals characterizes the time spacing across all rows of the time delay embedded matrix. As such, handling irregularly spaced data requires specification of a distinct \mathbf{L}_h and consequently, $\mathbf{W}_{h,GLLA}$, for each row of the time delay embedding matrix, $\mathbf{Y}_{h,\tau}^{(d)}$. Alternatively, missingness may be inserted at appropriate places to yield equally spaced observations in the time delay embedding matrix.

One characteristic of the GLLA to note is that when $d = 3$, the level estimates, i.e., $\hat{y}_h(t)$ in $\tilde{D} \hat{\mathbf{Y}}_{h,\tau}^{(d)}$ are simply observed manifest observations, $y_h(t)$, from the second column of the time delay embedded matrix in Equation 17. The corresponding derivatives are also computed based on raw observed data without additional smoothing. With $d > 3$, a larger window of neighboring observations is used to compute the local level and derivative estimates (Hu et al., 2014). The greater overlap between successive windows “smooths” the level and derivative estimates, much like a weighted moving average process.

Stage 1 Using the Generalized Orthogonal Local Derivative (GOLD) Approach

Deboeck (2010) noted that the GLLA yields level and derivative estimates that are correlated with each other. Thus, when these correlated level and derivative estimates are used in model fitting (Stage 2), parameter estimates may be greatly biased. To alleviate this problem, Deboeck proposed using an alternative weighting scheme, termed the Generalized Orthogonal Local Derivative (GOLD), that orthogonalizes the level and all the resultant derivative estimates. Implementation is facilitated by the R code in Deboeck (2010). In the GOLD, the matrix of level and derivative estimates, $\tilde{D} \hat{Y}_{h,\tau}^{(d)}$ is computed as

$$\tilde{D} \hat{Y}_{h,\tau}^{(d)} = Y_{h,\tau}^{(d)} W_{h,GOLD} = Y_{h,\tau}^{(d)} (D \Xi_h)' [(D \Xi_h)(D \Xi_h)']^{-1} \quad (21)$$

where D is a $(m_h + 1)$ square diagonal matrix with diagonal elements equal to $1/g!$, where $g = 0, 1, \dots, m_h$ and “ $g!$ ” denotes the factorial of g . Ξ_h is a $(m_h + 1) \times d$ matrix such that for $q = 0, 1, \dots, m_h$ and $r = 1, \dots, d$, the $(q+1$ th, r th) entry is given by

$$\Xi_{h,q+1,r} = t_r^q - I_{(q>0)} \left\{ \sum_{p=0}^{q-1} \Xi_{h,p+1,r} [(\sum_r \Xi_{h,p+1,r} t_r^q) / (\sum_r \Xi_{h,p+1,r} t_r^p)] \right\} \quad (22)$$

where t_r^q denotes the measurement time of the observations in column r raised to the q th power, and $I_{(q>0)}$ is an indicator function such that the operation after t_r^q is only performed for $q > 0$. We note that, as with the GLLA, the GOLD set-up assumes that the observations in the r th column of the time delay embedding matrix are separated by the same time interval relative to the observations in the $(r - 1)$ th column across all rows. In other words, while the observations may be unequally spaced within each row of the time delay embedding matrix, the same set of time intervals holds across all rows of the time delay embedding matrix. For data where the time intervals differ both within a row and across rows, the GOLD, similar to the GLLA, requires the specification of a distinct Ξ_h and consequently, $W_{h,GOLD}$, for each row of the time delay embedded matrix.

Stage 2: Fitting mixed effects ODE model with GLLA and GOLD

After the selection of d , the GLLA and GOLD use related but distinct matrices of predefined weights – $W_{h,GLLA}$ and $W_{h,GOLD}$, respectively – and a time delay embedded data matrix, to obtain derivative estimates. Elements in the columns of $\tilde{D} \hat{Y}_{h,\tau}^{(d)}$ are then used as observed variables for fitting a mixed effects model as in Stage 2 of FODEmixed. Specifically, for each observed time point, $t_{i,j}$, $\hat{x}_{h,i}^{(1)}(t_{i,j})$ is approximated by $\mathcal{Y}_h(t_{i,j})$; $\hat{x}_{h,i}^{(2)}$ is approximated by $\left[\frac{d\hat{y}_{h,i}(t_{i,j})}{dt} \dots \frac{d^{m_h-1}\hat{y}_{h,i}(t_{i,j})}{dt^{m_h-1}} \right]'$, namely, elements from the second to the m_h th column of $\tilde{D} \hat{Y}_{h,\tau}^{(d)}$, finally, $\frac{d^{m_h}\hat{y}_{h,i}(t_{i,j})}{dt^{m_h}}$, contained in the last column of $\tilde{D} \hat{Y}_{h,\tau}^{(d)}$ is used to approximate

the l th element in $\frac{d^{(m)}\hat{x}_i(t)}{dt^{(m)}}$. These elements are then used as observed variables to fit the model in Equation 16 using standard linear or nonlinear mixed effects estimation routines, exactly as described in the presentation of the FODEmixed approach above.

Similarities and Differences among FODEMixed, GLLA and GOLD

We note some conceptual similarities between the GLLA/GOLD method and the FDA-based method. Selection of τ in GLLA/GOLD is functionally similar to the placement of the knot points in the FDA-based method. In both approaches, some measurement occasions may be explicitly “skipped” or underweighted. When the underlying change process is oversampled, or the fluctuations between immediately adjacent measurement occasions reflect noise, setting $\tau > 1$ or placing knot points more sparsely across the time series contributes to “smoothing” of the level and derivative estimates. In the present context, we set $\tau = 1$ in the GLLA/GOLD and the knots points for the FDA to coincide with the observed time points. Thus, no observation is skipped and this aspect does not contribute to differences in the performances of the three methods.

The methods also differ with respect to whether the derivatives are computed based on smoothed vs. unsmoothed level data. In the GLLA/GOLD, the amount of data smoothing is determined solely by the order of d when $\tau = 1$. In the GLLA, using $m_h + 1$ weights in each row of the loading matrix, \mathbf{L}_h to compute derivative estimates has the effect of “approximating” the raw data using Taylor series expansion/polynomials of order m_h and using coefficients from such a Taylor series approximation to fix (i.e., not estimate) the values of the weight matrices \mathbf{L}_h and $\mathbf{\Xi}_h$ (see Equations 20; Hunter, accepted with revisions). The derivative estimates are thus computed using a window of raw (i.e., unsmoothed) data, although some smoothing can be expected by computing the derivatives from a window of neighboring points. In the GOLD, the weights from the Taylor series expansion are further subjected to linear transformations to orthogonalize the level and derivative estimates. Thus, derivative estimates from the GLLA, and to a lesser degree, GOLD, tend to be “rougher” compared to those from the FDA because derivatives in the latter are computed using smoothed level data (see Equation 15) weighted by means of the estimated basis coefficients, c_k , obtained by minimizing (13). The extent of smoothing in the FDA is further governed by the order of the B-spline, the choice and order of the penalty function in Equation 14, the smoothing parameter, λ_h and if so desired (but not done in the current context), the number and placement of knot points. Some differences between the GLLA and the GOLD have already been noted by Deboeck (2010) and are not reiterated here. Several other differences that have not been addressed in previous studies will be highlighted next and evaluated in a simulation study.

Unanswered Questions about Derivative Estimates from the FDA, GLLA and GOLD

Previous studies utilizing the GLLA and GOLD have focused exclusively on linear ODE models such as the damped oscillator model (Boker et al., 2010; Deboeck, 2010). Their performances in estimating derivatives from *nonlinear* ODE models relative to the FDA and each other are unknown. In addition, Deboeck (2010) considered very small embedding dimensions (e.g., $d = 3, 4$) and showed that the GOLD outperformed the GLLA in those

situations. The performances of these approaches in situations with very high d (e.g., $d > 10$ as considered in Boker et al. (2014) have not yet been evaluated. Using the van der Pol oscillator model in Equation 11 with $\zeta = 1$, $T = 200$ and $n = 1$, we will highlight here some of the distinct issues that may arise when derivative estimates are used to fit nonlinear mixed effects ODE models. Specifically, when nonlinear ODE models are involved, all three approaches, including the FDA, GLLA and GOLD, may yield reasonable derivative estimates when no measurement noise is present. For the GLLA or GOLD, a relatively low embedding dimension (e.g., $d = 5$) may be sufficient (see Figures 3(A)–(B)). However, when there is some measurement noise, such as when reliability = 0.9,⁶ derivative estimates from the GLLA and GOLD, particularly the second derivative estimates, were very noisy (see Figures 3(C)–(D)). In fact, to avoid cluttering the second derivatives plot in Figure 3(D), we restricted the range of the plot to $[-10, 10]$. The actual ranges of the derivative estimates from the two approaches at $d = 5$, as shown in the boxplots in the bottom panels of Figure 3, were much larger than the range of the true second derivatives.

The GLLA, however, does have one strength over the GOLD that was previously unexplored. In the GLLA, it is possible, in principle, to specify the extraction of derivatives of a higher-order than those required in a hypothesized ODE model. Even if these higher-order derivatives are not utilized in fitting the hypothesized mixed effects ODE, doing so has the effect of changing the order of the Taylor series polynomials used to approximate the level and subsequently, changes the weights that appear in L_h and by extension, the derivative estimates. The derivative estimates can be expected to be more “wiggly” in this case because they are constructed to capture more of the subtle deviations in the data. The weights in the GOLD, in contrast, are unaltered by the highest-order of the derivatives a researcher extracts because the weight matrix is designed to orthogonalize the relations among level and derivatives of all orders. When a 4th-order GLLA is coupled with a low embedding dimension of 5 to estimate up to second-order derivatives, the corresponding estimates are extremely noisy (see the boxplot in panel G of Figure 3). However, when the 4th-order GLLA is used with an extremely high embedding dimension – namely, 40, a number that is much higher than has been previously considered – the 4th-order GLLA yields second derivative estimates that appear comparable to those obtained using the FDA (see Figure 3(D)). However, doing so led to a time delay embedded matrix with length $T - (d - 1) = 161$ rows of data and consequently, a time series of second derivative estimates of length 161, in contrast to the full length of $T = 200$ second derivative estimates from the FDA.

Intrigued by these possibilities, we extend knowledge of FDA, GLLA and GOLD approaches for derivative estimation by examining their performances with (1) a nonlinear ODE; (2) a broader range of generally higher values of embedding dimensions than those utilized in previous studies that compared the GLLA and GOLD (e.g., $d = 3, 4$ used in Deboeck, 2010); and (4) both 2nd- and 4th-order specification in the GLLA.

⁶Note that the model-implied variance functions of a nonlinear ODE model generally do not have a closed analytic form. Thus, we computed the estimated reliability associated with a particular ODE model using simulated data generated from the model. Reliability was computed heuristically as the ratio between the observed variance of repeated realizations of the true scores and the observed total variance at the specified n and T .

Simulation Study

We conducted a simulation study to evaluate the performance of the proposed FODEmixed approach for fitting nonlinear ODE models with random effects, and additionally, to assess the relative strengths and weaknesses of the FODEmixed, GLLA and GOLD. To accomplish these goals, we considered a nonlinear, second order ODE model adapted from our empirical motivating example. Numerical solutions of the hypothesized ODE obtained using the Livermore numerical solver, *lsoda*, in the R library, *deSolve* (Hindmarsh, 1983; Petzold, 1983), were used as true values of the bivariate processes from $t_{i,1} = 0$ to $t_{i,T} = 20$ for all individuals, with four sample size configurations: (1) $T = 150$, $n = 100$, (2) $T = 150$, $n = 200$, (3) $T = 300$, $n = 100$, and (4) $T = 300$, $n = 200$.

The model considered was formulated as

$$\begin{aligned} \frac{d^2 x_{1i}(t)}{dt^2} &= -\eta_{1,i} x_{1i}(t) + a_{12} \frac{dx_{1i}(t)}{dt} x_{2i}(t) \\ \frac{d^2 x_{2i}(t)}{dt^2} &= -\eta_{2,i} x_{2i}(t) + a_{21} \frac{dx_{2i}(t)}{dt} x_{1i}(t) \\ \eta_{1,i} &= \eta_{10} + \eta_{11} u_{1i} + \eta_{12} u_{2i} + b_{\eta_{1,i}} \\ \eta_{2,i} &= \eta_{20} + \eta_{21} u_{2i} + \eta_{22} u_{2i} + b_{\eta_{2,i}} \end{aligned} \quad (23)$$

where $\eta_{1,i}$ and $\eta_{2,i}$ are person-specific parameters that govern the frequencies of the two processes. The terms $\frac{dx_{1i}(t)}{dt} x_{2i}(t)$ and $\frac{dx_{2i}(t)}{dt} x_{1i}(t)$ introduce nonlinearities into the model. Consonant with the theoretical properties of children's prepotent responses and executive process described in our motivating example, this model posits that the extent to which each process shows damping or amplification over time is moderated by the level of the other process. Articulating the idea that the executive process is invoked to override prepotent responses and that heightened prepotent responses may, in turn, diminish or "extinguish" a child's executive process, we expect bidirectional coupling wherein damping in each of the two processes is driven – or *moderated* – by the level of the other process. Conceptually, the model has some parallels to the classic Van der Pol oscillator model (see Equation 11) in which the damping in a univariate process is moderated by the squared level of the process, yielding especially pronounced damping in magnitude at extreme levels of the process. In our model, the damping is moderated by the level of an "extrinsic," opposing process that is competing for the same resources, rather than by the "intrinsic" level of the process. In cases where there is bidirectional coupling, whether the two processes show disproportionately high damping or amplification depends both on the signs of a_{12} and a_{21} , as well as whether the other process is above or below its equilibrium level (= 0 in this specific model).

The two person-specific parameters, $\eta_{1,i}$ and $\eta_{2,i}$, were assumed to conform to a linear combination of fixed and random effects. To improve the numerical stability of the estimation algorithm, we constrained the baseline frequency parameters, η_{10} and η_{20} , to be negative and optimized the log of $-\eta_{10}$ and $-\eta_{20}$ instead. u_{1i} and u_{2i} were fixed covariates: u_{1i} was assumed to conform to a uniform distribution between the range of 0 and 4, whereas u_{2i} was assumed to be a Bernoulli distributed binary variable with a .5 probability of obtaining a 1.0. The random effects vector, \mathbf{b}_i , consists of two random effects, $\mathbf{b}_i = [b_{\eta_{1,i}},$

$b_{\eta_2,i}$, and is assumed to be multivariate normally distribution with mean vector of zeros and standard deviations given by $\Sigma_b^{1/2} = \text{Diag} \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$. The other parameters used to simulate the data were set to the values of $\eta_{10} = 0.41$, $\eta_{20} = 0.00$, $\eta_{11} = -0.5$, $\eta_{12} = 0.0$, $\eta_{21} = 0.0$, $\eta_{22} = -0.3$, $a_{12} = -0.02$, $a_{21} = -0.03$. Measurement errors drawn independently from two normal distributions with means of zeros and SDs, $\sigma_{\varepsilon_1} = 1.0$ and $\sigma_{\varepsilon_2} = 1.0$, respectively, were added to $x_1(t_{i,j})$ and $x_2(t_{i,j})$ to obtain noisy observed measurements, $y_1(t_{i,j})$ and $y_2(t_{i,j})$. Using these parameters led to data with an average approximate reliabilities of .89 and .87, respectively. Reliability for each subject and process was calculated as in our earlier illustrations as the ratio between the variance of the simulated true scores and variance of the observed scores (with measurement errors) for each subject and process.

The FODEmixed approach was implemented as outlined in Equations 12–15 with the order of the B-spline set to 6. The smoothing parameter for each of the two processes, λ_b , was selected from the range of [1e-06, 10] and was set to be the value that minimized the average GCV value across individuals. The GLLA and GOLD were implemented with τ set to 1 and five different embedding dimensions: $d = 5$ (the minimum dimension needed for a 4th-order GLLA), 10, 30, 40, and 50. Preliminary examination of estimation results using the GLLA and GOLD with $d < 5$ suggested continued improvements in estimation results with $d > 5$ for both the GLLA and GOLD so we did not include $d < 5$ in our full Monte Carlo simulations.

To summarize, we considered 4 sample size configurations x 3 estimation approaches (FODEmixed, GLLA and GOLD) within which 5 possible embedding dimensions were used for the GLLA and GOLD, yielding a total of 44 conditions. Five hundred Monte Carlo (MC) replications were conducted for each condition, and the results were quantified in several ways. The root mean squared errors (RMSEs) and relative biases were used to summarize properties of the point estimates. The empirical SE of a parameter (i.e., standard deviation of the parameter estimates across all MC 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 the SE estimates (denoted as RDSE, namely, the difference between the average SE estimate and the true SE over the true SE). In addition, because the *nlme* routine in *R* does not provide SE estimates for all variance and covariance parameters, we also report the 95% confidence intervals for all parameters as averaged across the MC runs. We then computed power estimates for parameters whose true values were zero by tallying the proportion of MC trials in which the 95% CIs did not include zero. For parameters that had a true value of zero (i.e., η_{12} and η_{21}), this proportion can be taken as a type I error estimate. All iterations in which any of the parameter and SE estimates contained non-finite values or *nlme* did not converge in 200 iterations were tallied as non-convergent cases.

Simulation Results

Results from fitting the nonlinear ODE model are summarized in Figures 4(A)–(D), and detailed further in Tables 1–4. Accuracy of the point estimates from the FODEmixed, as indicated by relative biases and RMSEs, was reasonable given the nonlinearities in the model and the moderate sample sizes considered in the present study. Increasing the number of total time points from 150 to 300 provided notable improvements in the accuracy of the

point estimates (as evidenced by the lower RMSEs and relative biases of the parameters), and some gain in efficiency (as indicated by smaller MC SDs and SE estimates). Doubling the number of participants from 100 to 200 only had negligible effects on the accuracy of the point estimates, but more notable impact on improving the efficiency of both the fixed and particularly the random effects parameters (as shown by the smaller MC SDs). The point estimates for the random effect SDs showed good accuracy with $T \geq 150$ and as few as 100 participants.

Small biases in SEs in comparison to the MC SDs were observed for some of the fixed effects parameters. Inspection of the RDSE for each parameter as shown in Tables 1—4 indicated substantial underestimation of the SEs for the parameters a_{12} and a_{21} , but reasonable SE estimates for other parameters. Slight increases in the bias of the SE estimates was found with larger T , but the biases remained small except for a_{12} and a_{21} —the key sources of nonlinearities in the model. Power estimates for the parameters whose true values were significantly different from zero were satisfactory even in the smaller sample size configurations, and continued to improve with larger T and n . Power estimates were generally close to 1.0 when $n = 200$ and $T = 300$.⁷ Type-I error rates for the parameters whose true values were equal to zero, namely, η_{12} and η_{21} , were close to zero in all conditions.

The performances of the three approaches are summarized in Tables 5–7. Power estimates were consistently high for all approaches at the sample size configurations considered. On average, the FODEmixed as well as the other three approaches (at their optimal d) all yielded power estimates that were generally close to, or exceeded .9 across all parameters whose true values were not equal to 0 in the population. For RMSEs and relative biases of the point estimates, we averaged these measures across all fixed effects parameters (including all the parameters shown in Tables 1–4 except for all the standard deviation parameters marked with σ) and random effects parameters (including σ_{η_1} , σ_{η_2}) separately as in previous MC studies involving nonlinear ODEs (Chow et al., 2016, 1). Biases in SE estimates were only available for the fixed effects parameters as noted above. The simulation results indicated the “best” embedding dimension for the LLA-based approaches differed by approach, with the GOLD generally attaining better estimation results than the GLLA at lower embedding dimensions. That is, parallel to the results shown by Deboeck (2010, with $d = 3$ and 4 in the context of a linear oscillator model with no damping), we found that the GOLD always outperformed the GLLA at $d = 5$ and 10 (note that comparative results were not available from the 4th-order GLLA due to its poor convergence property at low d).⁸ However, the notable advantages of the GOLD relative to the GLLA were not present at the higher embedding dimensions. In fact, based on the plots in Figures 4A–D in which only results associated with the “best” embedding dimension for each of the LLA-based approaches are shown, the performances of the GOLD ($d = 10$) and the GLLA ($d = 30$) are similar.

⁷Note that this finding pertained also to the parameter $\log(\eta_{20})$, whose true value was 0 because $\eta_{20} = 1.0$.

⁸The 4th-order GLLA did not yield any converged cases in all the conditions with $d < 40$. We therefore report in Table 6 only the results from the conditions with $d = 40$ and 50.

Compared to the GLLA and GOLD, the FODEmixed was associated with lower RMSEs for the fixed as well as random effects parameters – an advantage that is more pronounced at larger T . Consistent with the detailed results shown in Tables 1–4, notably lower RMSEs were observed for the fixed effects parameters from FODEmixed with larger T , but the improvements with larger n were much smaller. RMSEs for the random effects SD parameters were closer between the FODEmixed and the two GLLA-based approaches, with the GOLD consistently showing distinctly higher RMSEs in random effects SDs across all sample size conditions (see Figure 4B). Given that all the approaches considered were two-stage approaches, some biases in the SE estimates are expected. Our simulation results indicated that biases in the SEs were indeed observed for all approaches (see Figure 4C), particularly for parameters associated with the nonlinear terms (i.e., a_{12} and a_{21}), but the biases were relatively small and averaged to zero across all fixed effects parameters (in other words, there were no systematic under- or over-estimation in SEs).

Interestingly, of the three LLA-based derivative approaches considered, the 4th-order GLLA, when used with a very large embedding dimension ($d = 50$), yielded estimation results similar to the FODEmixed in terms of RMSEs and also outperformed the 2nd-order GLLA as well as the GOLD. Specifically, while the FODEmixed yielded the lowest RMSEs for the point estimates, the 4th-order GLLA was, at times, characterized by slightly lower relative biases than the FODEmixed (see Table 6). The 4th-order GLLA was also associated with the smallest biases in SE estimates, but possibly because of all approaches, this approach was used with the smallest sample sizes when $d = 50$. However, the proportions of non-convergence were very high for the three LLA-based approaches in general with large n , with the proportions of non-convergence being greater than 80% in the largest sample size condition. Increasing the maximum number of iterations from the default 50 iterations to 200 iterations still did not yield more convergent cases. The convergence problems were especially severe in the 4th-order GLLA conditions with low embedding dimensions (i.e., < 50), wherein most replications simply did not converge and the few cases that did converge were characterized by very biased estimates. It is speculative, but individual differences in the dynamic parameters might have given rise to a wider repertoire of distinct change trajectories with large n and T that requires further tuning and explorations to identify the “optimal” embedding dimension. Another possibility is that biases in some of the dynamic parameter estimates might have led to additional correlations among the random effects. Forcing the corresponding random effects to be uncorrelated to mirror the true model might have aggravated the convergence problems as well, especially with increased n .

Finally, while the GOLD, just as the FODEmixed, appeared consistent (i.e., the empirical sampling distributions of the parameter estimates approached the true population values with increased sample sizes, particularly with larger T), the quality of the estimates from the two GLLA approaches appeared to depend more closely on the choice of d , rather than increases in T (see Tables 5–7).

Empirical Illustration

Observations of children’s and mothers’ behaviors during an 8-minute wait task were used in this illustration. During the wait task (Cole et al., 2011; Vaughn, Kopp, & Krakow, 1984),

children were told that they had to wait to open a gift until their mothers completed work. The delay in opening the gift, the absence of interesting alternative activities, and the loss of the mother's attention jointly created a frustrating situation for the children. We used composite scores of prepotent responses (PR) and executive process (EP) obtained from aggregating 3 seconds of video-coded data over the course of 8 minutes, yielding approximately $T = 160$ measurement occasions for each of $n = 115$ participants. Behaviors coded as PR included children's anger intensity, angry bids to mother, focusing on gift, touching the gift and other disruptive behaviors. Indices of EP included rankings of strategic behaviors with 1 = self-soothing and 5 = focused distraction, with higher ranks assigned to strategies that were of higher developmental maturity and decreased reliance on external sources of regulation (e.g., the mother). A plot of the EP and PR scores from 5 randomly selected individuals is shown in Figure 5(A).

Two person-specific covariates were used. First, we included child effortful control, assessed using the Toddler Behavior Assessment Questionnaire (TBAQ-R; Goldsmith, 1996). Effortful control is a temperament dimension that reflects the degree to which children are disposed to modulating their attention, behavior and emotion (Rothbart & Bates, 2006), and is thought to be a child's contribution to the development of self-regulated behavior. Second, we included mothers' structuring quality – a parenting dimension involving attempts to foster a child's self-regulation (Hoffman, Crnic, & Baker, 2006) – as the degree of sensitive, appropriate attempts by the mother to help the child tolerate the wait (0 = no structuring to 3 = highly sensitive structuring).

We used the FDA approach outlined in Equations 12–15 with B-splines of order 6 to obtain smoothed estimates of the participants' levels, first derivatives and second derivatives of EP and PR scores. The knot points were specified to coincide with the observed time points, and the integrated square of the fourth derivative was used in the penalty function in Equation 14. The smoothing parameter, λ_h , was chosen separately for each individual using the GCV. Each individual's PR and EP time-series were demeaned (mean-centered) so that the resultant EP and PR scores captured the within-person deviations in EP and PR. Throughout, we use $EP_i(t)$ and $PR_i(t)$ to denote the i th child's demeaned levels EP and PR at time t , respectively; $\frac{dEP_i(t)}{dt}$ and $\frac{dPR_i(t)}{dt}$ denote the two processes' first derivatives at time t , and $\frac{d^2EP_i(t)}{dt^2}$ and $\frac{d^2PR_i(t)}{dt^2}$ denote the second derivatives at time t .

To probe for possible presence of nonlinearities, we fit a regression model separately to the participants' demeaned PR and EP scores, in which the linear effects of the levels and first derivatives of each process, as well as the levels and first derivatives of the “opposing process,” were included as predictors of the second derivatives in each process. The component + residuals plots (see a selected subset in Figure 5(B)–(E)) were used to identify associations between the second derivatives of a process and other terms after the effects of the linear terms explicated above have been accounted for. For most participants, the plots depicted a negative relation between the PR and EP deviation scores and their second derivatives, thus suggesting the presence of oscillatory patterns. There was evidence that this relation was nonlinear, particularly for PR (i.e., the Loess line deviated more from the linear least squares line in Figure 5(B) compared to Figure 5(C)). We then explored the partial

correlations between the residuals of the linear model and additional nonlinear predictor terms, including (1) the quadratic level terms, $PR_i(t)^2$ and $EP_i(t)^2$; (2) the quadratic first derivative terms, $(dPR_i(t)/dt)^2$ and $(dEP_i(t)/dt)^2$, based on Figure 5(E); and (3) product terms such as $PR_i(t)EP_i(t)$, $(dEP_i(t)/dt)PR_i(t)$, $(dPR_i(t)/dt)EP_i(t)$, $(dEP_i(t)/dt)PR_i(t)^2$ and $(dPR_i(t)/dt)EP_i(t)^2$. Terms with substantial partial correlations were incorporated into the final nonlinear mixed effects model. All estimation in Stage 2 was performed using the nlme package in R with restricted maximum likelihood.

Based on results from the preliminary model explorations and our theoretical beliefs, we first considered the full model:

$$\begin{aligned} \frac{d^2 EP_i(t)}{dt^2} = & \eta_{EP} EP_i(t) + \zeta_{EP,1} \left(\frac{dEP_i(t)}{dt} \right) + \zeta_{EP,2} \left(\frac{dEP_i(t)}{dt} EP_i(t)^2 \right) + \gamma_{EP,1} PR_i(t) \\ & + \gamma_{EP,2} \left(\frac{dPR_i(t)}{dt} \right) + \gamma_{EP,3} (EP_i(t) PR_i(t)) + \gamma_{EP,4} \left(\frac{dEP_i(t)}{dt} PR_i(t)^2 \right) \\ \frac{d^2 PR_i(t)}{dt^2} = & \eta_{PR} PR_i(t) + \zeta_{PR,1} \left(\frac{dPR_i(t)}{dt} \right) + \zeta_{PR,2} \left(\frac{dPR_i(t)}{dt} PR_i(t)^2 \right) + \gamma_{PR,1} EP_i(t) \\ & + \gamma_{PR,2} \left(\frac{dEP_i(t)}{dt} \right) + \gamma_{PR,3} (EP_i(t) PR_i(t)) + \gamma_{PR,4} \left(\frac{dPR_i(t)}{dt} EP_i(t)^2 \right) \end{aligned} \quad (24)$$

Some of the parameters in this model (e.g., η_{EP} , η_{PR} , $\zeta_{EP,1}$ and $\zeta_{PR,1}$) are parameters that also appear in the linear oscillator model in Equation (8). Exploratory evidence suggested the need to allow the quadratic trends seen in Figures 5 (B)–(E) to be moderated by other level and derivative variables. Consequently, we included the terms $\left(\frac{dEP_i(t)}{dt} EP_i(t)^2 \right)$ and $\left(\frac{dPR_i(t)}{dt} PR_i(t)^2 \right)$. Their corresponding regression slopes, $\zeta_{EP,2}$ and $\zeta_{PR,2}$, may be interpreted as the “intrinsic regulatory” parameters for EP and PR, respectively, in that they capture the additional damping (if negative) or amplification (if positive) in these processes when their levels deviated far from the person-specific EP and PR means (a characteristic of the van der Pol system in Equation 11).

Four kinds of coupling/moderating influences between EP and PR were considered. The first two kinds were linear coupling effects from the opposing process’ level, (as reflected in

$\gamma_{EP,1} PR_i(t)$ and $\gamma_{PR,1} EP_i(t)$) and first derivative (as captured by $\gamma_{EP,2} \left(\frac{dPR_i(t)}{dt} \right)$ and $\gamma_{PR,2} \left(\frac{dEP_i(t)}{dt} \right)$) had been considered previously in other work featuring coupled linear oscillators (e.g., Hu et al., 2014). These terms capture the effects of how the curvatures or “accelerations”/“decelerations” in each process are related to the level and instantaneous changes in the opposing process. The third kind of coupling effects, reflected in the coupling coefficients $\gamma_{EP,3}$ and $\gamma_{PR,3}$, is linked to the product term, $(EP_i(t) PR_i(t))$. These interactive level influences capture how EP and PR may induce added fluctuations in the other process as they “interact and compete” to be the dominant mode of behavior. These inter-process interaction effects are conceptually similar to how predator-prey interactions influence population growth and decline in the classical predator-prey model in Equation 10. However, here these inter-process interactions influence curvatures (second derivatives) rather than rates of change (first derivatives). Finally, the coefficients $\gamma_{EP,4}$ and $\gamma_{PR,4}$, tied to the product

terms $\left(\frac{dEP_i(t)}{dt}PR_i(t)^2\right)$ and $\left(\frac{dPR_i(t)}{dt}EP_i(t)^2\right)$, respectively, were motivated by our theoretical model of how the damping and amplification in PR and EP might be moderated by the levels of the opposing process (as discussed in the *Motivating Example* section).⁹ This characteristic is similar to the two intrinsic regulatory parameters explained above ($\zeta_{EP,2}$ and $\zeta_{PR,2}$), except that now damping (or alternatively, amplification) is induced by the extrinsic opposing process.

We then explored the need to allow for random effects in subsets of the parameters shown in Equation 24 by sequentially adding/omitting random effects and evaluating whether the 95% CIs for their SDs included zero. Examination of the 95% CIs and follow-up residual plots revealed considerable between-individual heterogeneities in multiple aspects of individuals' dynamics and the need to include random effects for almost all dynamic parameters. For exploratory purposes, we retained most of the of the random effect terms (a total of 10) to gain insights into the extent and nature of such interindividual differences (see Table 8; parameters with random effects are marked with a subscript i).¹⁰ Given our interest in exploring correlates of individual differences in the coupling and intrinsic regulatory parameters, effortful control and maternal structuring quality were added as predictors of individual differences in these parameters. We then sequentially removed fixed effects terms that were not statistically significant and parameter estimates from the “final” model are shown in Table 8.

The intercepts of 5 of the 8 coupling parameters considered were statistically different from zero. Significant interindividual differences were found for all coupling parameters and the frequency-related parameters, $\eta_{EP,1,i}$ and $\eta_{PR,1,i}$. To illustrate the effects of the estimated parameters on the participants' EP and PR dynamics, we plotted the observed and predicted (model-implied) EP and PR trajectories of one randomly selected individual (see Figure 6(A)) using the parameter estimates from the final model and the observed initial conditions of the participant. The plot suggested that the model was able to capture some of the more pronounced curvatures in the observed data, but some evidence of misfit can also be observed (as we discuss below).

The intercepts of $\zeta_{EP,1,i}$ and $\zeta_{PR,1,i}$ representing the linear effect of each process' first derivative on its second derivative for individuals with average effortful control and maternal structuring quality, were not statistically significant. The intercepts of $\zeta_{EP,2,i}$ and $\zeta_{PR,2,i}$ which represent the two process' intrinsic regulatory strengths in the “average” individual, were also not statistically significant. Substantial individual differences were found in all of these parameters, and effortful control was found to have a significant negative effect on $\zeta_{EP,2}$. As depicted in Figure 6(B), the EP of a hypothetical child with high effortful control (+2SD above the sample mean), when compared to a child with low effortful control (−2SD

⁹We initially expected the moderating effect of the opposing process on the damping/amplification of a process to originate from the level of the opposing process (as captured by the parameters a_{12} and a_{21} in the simulation model; see Equation (23)). Evidence from preliminary model explorations suggested the need to consider, instead, the quadratic moderating effect of the level of the opposing process.

¹⁰Throughout, we limited ourselves to a diagonal variance structure for the random effects because the approximate Hessians for the parameters were not positive definite and the 95% CIs for the parameters could not be obtained when correlations were allowed among the random effects, even with as few as three random effects in the model.

below the sample mean), returned to zero (the person-specific mean) more quickly at extreme values of EP. This intrinsic EP characteristic of children higher in effortful control can be interpreted as more constancy in the use of EP-related strategies. As the ability to initiate and sustain strategic attempts in frustrating situations is still nascent at 36 months of age, this finding represents a positive developmental change. In contrast, the higher variability shown by the low effortful control children may reflect these children's haphazard attempts at using a variety of EP strategies to regulate their PR.

A statistically significant negative effect of the level of PR on the second derivative of EP ($\gamma_{EP,1}$) was found, and higher effortful control was associated with more negative values on this parameter (see Figure 6(C)). The negative value of $\gamma_{EP,1}$ suggested that EP tended to accelerate when PR was low (i.e., negative or below the person's mean EP) and decelerate when the child's PR was high (i.e., positive or above the person's mean PR). Thus, for children high in effortful control, EP accelerated more and "prevailed" longer when PR was low. In contrast, when PR was high, the child's EP also decelerated (slowed down) more and "turned around" more quickly to show rises again. Overall, the magnitude and sign of the intercept of $\gamma_{EP,1}$ were consistent with the role of effortful control assumed in the literature: when PR is low, it is easier to engage developmentally mature EP strategies, particularly for the high effortful control children; however, when PR is high, children with high effortful control show a quicker reversion to more mature EP strategies than children with low effortful control.

Significant nonlinear interactive effects were also found in the group as a whole, with the intercepts for the coupling parameters, $\gamma_{EP,3,i}$, $\gamma_{PR,3,i}$ and $\gamma_{PR,4,i}$ all being significantly different from zero. The negative estimated intercept value of $\gamma_{EP,3,i}$ and positive intercept value of $\gamma_{PR,3,i}$ suggested that when these processes were in synchrony (i.e., both high or both low), the interaction between the two processes tended to lead to deceleration (or "slowing down") in the changes in EP but acceleration ("speeding up") in the changes in PR. In contrast, when they were asynchronous (i.e., one was high when the other was low), the reverse was observed. Thus, when EP was high (positive) and PR was low (negative), as in Figure 6(D) around $t = 15$, $EP_{\lambda}(t)PR_{\lambda}(t)$ became negative and the resultant interaction boosted the level of EP (due to the negative value of the intercept of $\gamma_{EP,3,i}$) while simultaneously slowing down the rise in PR (due to the positive value of the intercept of $\gamma_{PR,3,i}$). In contrast, at around $t = 40$, a high (positive) level of EP was observed to "interact and compete with" a high (positive) level of PR to "speed up" or sustain the growth in PR while simultaneously slowing EP's return to 0 following a decline. Taken together, such interactive coupling effects suggested the tendency for PR to "thrive and dominate" when both PR and EP were high, and for EP to show accelerated changes only when the levels of EP and PR diverged from each other.

We also found a statistically significant EP-driven damping effect in PR, but not a PR-driven damping effect in EP, in the group as a whole (i.e., the intercept of $\gamma_{PR,4,i}$ but not $\gamma_{EP,4,i}$ was negative and significantly different from zero). This parameter has the same effect on PR as $\zeta_{2,EP}$ on EP, only that now PR was driven to damp toward zero by EP, as opposed to its own levels. The parameter estimate suggested that for this age group, PR was found to show greater damping (faster return to 0 or each child's person-specific PR mean) at extreme

values of EP. Thus, model explorations and subsequent fitting using the FODEmixed provided evidence for what may be termed *effective self-regulation* – that at 36 months old, children’s executive efforts begin to damp or override their prepotent responses over the course of a frustrating situation. Although effortful control and maternal structuring quality were not found to be significant predictors of this EP-induced damping effect, significant individual differences were found.

For comparisons, we fit the linear oscillator model with only linear coupling effects (i.e., the effects $\gamma_{EP,1}$ and $\gamma_{PR,1}$ in Equation 24) and random effects for all parameters to the data. The final nonlinear model was characterized by lower Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and estimated residual error SDs (AIC = 22345; BIC = 22564; $\sigma_{\varepsilon_{PR}}^* = 0.37$; $\sigma_{\varepsilon_{EP}}^* = 0.29$) than the linear comparison model with only statistically significant effects retained (AIC = 24888; BIC = 24981; $\sigma_{\varepsilon_{PR}}^* = 0.38$; $\sigma_{\varepsilon_{EP}}^* = 0.31$). Plots of the *discrepancies* between the observed and predicted trajectories from the nonlinear and linear models, shown respectively in Figures 6(E) and (F), revealed that the nonlinear model captured the peaks and valleys in the two processes (particularly in PR) more accurately (i.e., with smaller discrepancies) than the linear comparison model. Of course, the nonlinear model is by no means perfect, and several sources of misfit can be noted from Figure 6(A). For instance, the nonlinear model hypothesized that the two processes would show damping toward zero but in practice, these processes might approach a negative asymptote that was below the person-specific mean. In addition, some model misfit was to be expected given that the EP and PR scores were essentially averages of count data that were not symmetrically distributed around their means. Extending the proposed ODE model to allow for discrete data is an important extension in the future. Stochastic extensions of the proposed ODE would help incorporate some uncertainties into the prediction process, thereby allowing the processes to deviate more from their underlying drift functions. Allowing the parameters in the ODE model to vary over time is another possible extension. These alternative modeling options should be explored thoroughly in parallel with investigations of other sources of between-person heterogeneities.

Discussion

In the present article, we presented, tested, and illustrated the utility of FODEmixed, a two-stage approach for fitting linear or nonlinear ODE models with random effects and unknown initial conditions. Our simulations with a nonlinear oscillator model indicated that the proposed technique yields satisfactory point and SE estimates of most parameters. However, we remain cautious about the accuracy of the SE estimates of the dynamic parameters, especially for parameters that enter the ODEs nonlinearly, and in situations where the data may not be of sufficient length and time density to capture the complexity of the dynamic processes of interest (i.e., time-series with $T < 150$).

We also examined the relative strengths and limitations of two additional two-stage approaches, GLLA and GOLD, for estimation of nonlinear ODEs. Three main findings emerged. First, extending previous studies that showed that the GOLD outperforms the GLLA at low embedding dimensions, we found that the advantages of GOLD do not extend

to situations with higher embedding dimensions. The implication is that the GOLD approach may be useful for modeling situations where embedding dimension must be kept low (e.g., when there are only a small number of repeated measures). Across conditions, the FODEmixed consistently outperformed the GOLD and 2nd-order GLLA in terms of accuracy and efficiency of the point estimates. As well, we found that a novel 4th-order GLLA that uses higher-order information from the time series for derivative estimation may have some advantages when the time-series are long and it is practical to use a high embedding dimension. However, similar to the GLLA and GOLD, the convergence rates of the 4th-order GLLA were not satisfactory in conditions with a larger number of participants, and its convergence rates were much worse than the 2nd-order GLLA and GOLD at low embedding dimensions. Reasons for the lower converge rates require more thorough evaluations, but alternative ways of further stabilizing the 4th-order GLLA to maximize its strengths (e.g., by utilizing the fourth-order derivatives in Stage 2 with specialized constraints; Boker, 2007) are important to consider as we venture into the realm of nonlinear ODE modeling.

As a whole, results from this article prompt discussion of how the strengths of each of the three approaches may be integrated to obtain even better estimation. First, the cross-validation indices used in the FDA framework may serve as helpful, “model-free” tools in the selection of tuning parameters (e.g., d) in the GLLA and GOLD. Second, the FDA-based approach uses a derivative-based penalty term to regularize derivative estimates so that they are smooth. It is likely that such penalty terms can also be integrated into the GLLA/GOLD procedures as a way to reduce noise. Admittedly, other issues may arise when smoothing techniques are used (e.g., over- and under-smoothing); these issues apply to all three approaches and warrant caution in empirical applications of any of these techniques. Third, in the FDA framework, many candidate basis functions (e.g., Fourier basis function, wavelet basis functions) are available for approximating diverse forms of change (Ramsay & Silverman, 2005). These alternative basis functions may also be used in the GLLA and GOLD to capture more complex changes (e.g., discontinuous changes, changes that unfold across multiple time scales). Fourth, consistent with previous results (von Oertzen & Boker, 2013), the use of time-delay embedding has been shown to improve estimation results in the GLLA and GOLD approaches. It may be possible to further improve the estimation properties in the FDA framework by incorporating time-delay embedding procedures.

In Stage 2, the FODEmixed, GLLA and GOLD approaches all make uses of standard mixed effects framework to model interindividual differences in intraindividual change. This part of the procedures can also be extended in a number of ways. For example, latent class and regime-switching ODE models may be developed to investigate the existence of sub-populations and/or within-person phases with distinct dynamics. Arguments abound for why homogeneity of within-person dynamics may not hold in the study of human dynamics (Molenaar, 2004). To this end, local influence analysis (Cook, 1986; Zhu, Ibrahim, Lee, & Zhang, 2007) or other alternative outlier detection/person-specific procedures may be merged with the procedures outlined here to identify individuals whose dynamics may deviate substantially from others in the sample (e.g., Chow, Hamaker, & Allaire, 2009; Gates & Molenaar, 2012; Nesselroade & Molenaar, 1999; Ram et al., 2014). In addition, in both the didactic illustration and empirical example, we used a variety of regression diagnostics

and fit indices to evaluate and build the models. It is unclear whether fit indices such as the information criterion measures are appropriately sensitive to model misspecifications in two-stage procedures.

We did not consider the use of two-stage approaches for handling SDEs in the present article because fitting and diagnosing SDEs are characterized by added challenges that are yet to be resolved in the approaches considered. For instance, in contrast to ODE processes that are contaminated only with measurement noise, it is unclear whether the smoothing options available from standard FDA procedures can adequately distill the process as well as measurement noises in SDEs to reveal the true relations among the derivative variables. In addition, as Oud (2007) has shown, when the true underlying model is an SDE, fitting the ODE variation of the model using two-stage (specifically, LLA) procedures generally leads to high biases and misleading results. Thus, if the true underlying model is in fact an SDE (either with or without random effects), appropriate estimation procedures have to be used in the second stage to incorporate the correct process noise structure. The nonlinear mixed effects estimation routine used in Stage 2 of all the two-stage procedures considered in this article requires further adaptations to fulfill this role. Despite these challenges, SDEs capture uncertainties and complexities in change processes in ways that ODEs cannot, and this extension certainly warrants further investigation in the future.

Using the FODEmixed, we were able to develop a preliminary mixed effects ODE model that builds on and further extends conceptual models positing transitional shifts in self-regulatory ability at about 3 years of age. Of note, although not all the details were written out in the empirical illustration section, multiple models were rejected and retained for further consideration based on interpretability and fit. Our general conclusion was that, indeed, many aspects of self-regulation are not represented in standard linear models of change, and that at least some of those aspects can be described empirically using nonlinear ODE models.

In promoting the move into nonlinear models, we are often faced with concerns as to whether behavioral science data can support these explorations. While the sample size conditions considered in our simulation study may seem dense, there are many study designs that produce such data. In the empirical example, for instance, we used a very standard design wherein second-by-second coding of an 8-minute laboratory task were aggregated into 3-second epochs. At faster time-scales, studies involving physiological data typically collect time series spanning many thousands of occasions. At slower time-scales, our own work with experiencing sampling designs has obtained 150+ time-series of between 200 and 800 occasions each (Ram et al., 2014). Generally, advances in mobile and web technology, miniaturization of sensors, and widespread adoption of digital communication platforms are providing a fast-expanding repertoire of data suited for dynamic modeling (Ram & Grimm, 2015), including the methods described here.

Conclusion

Across the domains of science, many theoretical frameworks have been developed to explain if, how, and when entities change. In behavioral science, our conceptualizations of change

phenomena often involve complex feedback loops, discontinuities, and complex interdependencies that require nonlinear models. Unfortunately, the hurdles between us and the world of nonlinear mathematical models have been many. Our intent here has been to remove some of those hurdles by presenting and illustrating an accessible approach that facilitates mapping of empirical data to nonlinear dynamic models – both in helping to identify viable models, and also to reject those that are not useful. We look forward to what emerges in the discovery, description and understanding of nonlinear human dynamics.

Acknowledgments

Funding for this study was provided by NSF grant SES-1357666, NIH grants R01MH61388, R01HD07699, R01GM105004, Penn State Quantitative Social Sciences Initiative and UL TR000127 from the National Center for Advancing Translational Sciences.

References

- Arminger, G. Linear stochastic differential equation models for panel data with unobserved variables. In: Tuma, N., editor. *Sociological methodology* 1986. San Francisco: Jossey-Bass; 1986. p. 187-212.
- Banks, RB. *Growth and diffusion phenomena: mathematical frameworks and applications*. New York: Springer; 1994.
- Beskos A, Papaspiliopoulos O, Roberts G. Monte carlo maximum likelihood estimation for discretely observed diffusion processes. *The Annals of Statistics*. 2009; 37(1):223–245.
- Boker, S.; Laurenceau, J-P. Dynamical systems modeling: an application to the regulation of intimacy and disclosure in marriage. In: Walls, T.; Schafer, J., editors. *Models for intensive longitudinal data*. New York: Oxford University Press; 2006. p. 195-218.
- Boker, SM. Specifying latent differential equations models. In: Boker, SM.; Wenger, MJ., editors. *Data analytic techniques for dynamical systems*. Mahwah, NJ: Lawrence Erlbaum Associates; 2007. p. 131-159.
- Boker, SM.; Deboeck, PR.; Edler, C.; Keel, PK. Generalized local linear approximation of derivatives from time series. In: Chow, S.; Ferrer, E.; Hsieh, F., editors. *Statistical methods for modeling human dynamics: An interdisciplinary dialogue*. New York, NY, USA: Taylor & Francis; 2010. p. 161-178.
- Boker SM, Graham J. A dynamical systems analysis of adolescent substance abuse. *Multivariate Behavioral Research*. 1998; 33:479–507. DOI: 10.1207/s15327906mbr3304_3 [PubMed: 26753826]
- Boker, SM.; Neale, MC.; Klump, KL. A differential equations model for the ovarian hormone cycle. In: Molenaar, PCM.; Newell, K.; Lerner, R., editors. *Handbook of relational developmental systems: Emerging methods and concepts*. New York, NY, USA: Guilford Press; 2014. p. 369-391.
- Boker, SM.; Neale, MC.; Rausch, J. Latent differential equation modeling with multivariate multi-occasion indicators. In: van Montfort, K.; Oud, H.; Satorra, A., editors. *Recent developments on structural equation models: theory and applications*. Amsterdam: Kluwer; 2008. p. 151-174.
- Brown EN, Luithardt H. Statistical model building and model criticism for human circadian data. *Journal of Biological Rhythms*. 1999; 14:609–616. DOI: 10.1177/074873099129000975 [PubMed: 10643759]
- Chow SM, Lu Z, Sherwood A, Zhu H. Fitting nonlinear ordinary differential equation models with random effects and unknown initial conditions using the Stochastic Approximation Expectation Maximization (SAEM) algorithm. *Psychometrika*. 2016; 81:102–134. DOI: 10.1007/s11336-014-9431-z
- Chow S-M, Ferrer E, Nesselrode JR. An unscented kalman filter approach to the estimation of nonlinear dynamical systems models. *Multivariate Behavioral Research*. 2007; 42(2):283–321. [PubMed: 26765489]

- Chow SM, Hamaker EJ, Allaire JC. Using innovative outliers to detecting discrete shifts in dynamics in group-based state-space models. *Multivariate Behavioral Research*. 2009; 44:465–496. DOI: 10.1080/00273170903103324 [PubMed: 26735593]
- Chow SM, Nesselrode JR. General slowing or decreased inhibition? Mathematical models of age differences in cognitive functioning. *Journals of Gerontology Series B: Psychological Sciences & Social Sciences*. 2004; 59B(3):101–109. DOI: 10.1093/geronb/59.3.P101
- Cole PM, Martin SE, Dennis TA. Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development*. 2004; 75:317–333. DOI: 10.1111/j.1467-8624.2004.00673.x [PubMed: 15056186]
- Cole PM, Tan PZ, Hall SE, Zhang Y, Crnic KA, Blair CB, et al. Developmental changes in anger expression and attention focus: Learning to wait. *Developmental Psychology*. 2011; 47:1078–1089. DOI: 10.1037/a0023813 [PubMed: 21639619]
- Coleman, JS. The mathematical study of change. In: Blalock, HM., Jr; Blalock, A., editors. *Methodology in social research*. New York: McGraw-Hill; 1968. p. 428-478.
- Cook D. Assessment of local influence (with discussion). *Journal of Royal Statistical Society: Series B (Methodological)*. 1986; 48:133–169.
- Cronbach LJ, Furby L. How should we measure “change” - or should we? *Psychological Bulletin*. 1970; 74(1):68–80. DOI: 10.1037/h0029382
- Cudeck R, Klebe KJ. Multiphase mixed-effects models for repeated measures data. *Psychological Methods*. 2002; 7(1):41–6. DOI: 10.1037/1082-989X.7.1.41 [PubMed: 11928890]
- De Boor C. A package for calculating with B-splines. *SIAM Journal of Numerical Analysis*. 1977; 14:441–472. DOI: 10.1137/0714026
- De Boor, C. *A practical guide to splines*. New York: Springer-Verlag; 1978.
- Deboeck PR. Estimating dynamical systems: Derivative estimation hints from Sir Ronald A. Fisher. *Multivariate Behavioral Research*. 2010; 45:725–745. DOI: 10.1080/00273171.2010.498294 [PubMed: 26735716]
- Dierckx, P. *Curve and surface fitting with splines*. New York, NY, USA: Oxford Science Publications; 1993.
- Eilers P, Marx B. Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). *Statistical Science*. 1996; 11(2):89–121.
- Faraway, JJ. *Linear models with R*. 1. Boca Raton, FL: Chapman & Hall/CRC; 2004.
- Ferrer E, Helm JL. Dynamical systems modeling of physiological coregulation in dyadic interactions. *International Journal of Psychophysiology*. 2013; 88(3):296–308. DOI: 10.1016/j.ijpsycho.2012.10.013 [PubMed: 23107993]
- Fox, J. *Applied regression analysis and generalized linear models*. 2. New York: Sage; 2009.
- Gates KM, Molenaar PCM. Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *Neuroimage*. 2012; 63:310–319. DOI: 10.1016/j.neuroimage.2012.06.026 [PubMed: 22732562]
- Goldsmith HH. Studying temperament via construction of the toddler behavior assessment questionnaire. *Child Development*. 1996; 67:218–235. DOI: 10.1111/j.1467-8624.1996.tb01730.x [PubMed: 8605830]
- Gottman, JM.; Murray, JD.; Swanson, CC.; Tyson, R.; Swanson, KR., editors. *The mathematics of marriage: Dynamic nonlinear models*. Cambridge, MA: MIT Press; 2002.
- Hamerle A, Nagl W, Singer H. Problems with the estimation of stochastic differential equation using structural equations models. *Journal of Mathematical Sociology*. 1990; 16(3):201–220.
- Harrell, FEJ. *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
- Harris, CW., editor. *Problems in measuring change*. Madison, WI: University of Wisconsin Press; 1963.
- Hawkey LC, Burleson MH, Berntson GG, Cacioppo JT. Stress in everyday life: Cardiovascular activity, psychosocial context and health behaviors. *Journal of Personality and Social Psychology: Personality and Individual Differences*. 2003; 85(1):105–120.

- Hindmarsh, AC. ODEPACK, a systematized collection of ODE solvers. In: Stepleman, R., et al., editors. Scientific computing. Amsterdam: North-Holland; 1983. p. 55-64.
- Hoffman C, Crnic KA, Baker JK. Maternal depression and child functioning: Implications for children's emergent emotion regulation and behavioral functioning. *Parenting: Science and Practice*. 2006; 6:271–295. DOI: 10.1207/s15327922par0604_1
- Hu Y, Boker S, Neale M, Klump KL. Coupled latent differential equation with moderators: Simulation and application. *Psychological Methods*. 2014; 19(1):56–71. DOI: 10.1037/a0032476 [PubMed: 23646992]
- Hunter M. As good as GOLD: Gram-Schmidt orthogonalization by another name. *Psychometrika*. accepted with revisions.
- Jones, RH. Longitudinal data with serial correlation: A state-space approach. Boca Raton, FL: Chapman & Hall/CRC; 1993.
- Kaplan, D.; Glass, L. Understanding nonlinear dynamics. New York: Springer-Verlag; 1995.
- Liang H, Wu H. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of American Statistical Association*. 2008; 103(484):1570–1583. DOI: 10.1198/016214508000000797
- Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990; 46(3):673–687. DOI: 10.2307/2532087 [PubMed: 2242409]
- Lotka, AJ. Elements of physical biology. Baltimore: Williams and Wilkins; 1925.
- Lu ZH, Chow SM, Sherwood A, Zhu H. Bayesian analysis of ambulatory cardiovascular dynamics with application to irregularly spaced sparse data. *Annals of Applied Statistics*. 2015; 9:1601–1620. DOI: 10.1214/15-AOAS846 [PubMed: 26941885]
- MacCallum RC, Kim C, Malarkey WB, Kiecolt-Glaser JK. Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*. 1997; 32:215–253. DOI: 10.1207/s15327906mbr3203_1 [PubMed: 26761610]
- Madhyastha TM, Hamaker EL, Gottman JM. Investigating spousal in-fluence using moment-to-moment affect data from marital conflict. *Journal of Family Psychology*. 2011; 25:292–300. DOI: 10.1037/a0023028 [PubMed: 21480708]
- Mbalawata IS, Särkkä S, Haario H. Parameter estimation in stochastic differential equations with markov chain monte carlo and non-linear kalman filtering. *Computational Statistics*. 2013; 28(3): 1195–1223. DOI: 10.1007/s00180-012-0352-y
- Meredith W, Tisak J. Latent curve analysis. *Psychometrika*. 1990; 55:107–122.
- Merrilees CE, Goeke-Morey M, Cummings EM. Do event-contingent diaries about marital conflict change marital interactions? *Behaviour Research and Therapy*. 2008; 46:253–262. [PubMed: 18177629]
- Molenaar PCM. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology-this time forever. *Measurement: Interdisciplinary Research and Perspectives*. 2004; 2:201–218. DOI: 10.1207/s15366359mea0204_1
- Molenaar PCM, Newell KM. Direct fit of a theoretical model of phase transition in oscillatory finger motions. *British Journal of Mathematical and Statistical Psychology*. 2003; 56:199–214. DOI: 10.1348/000711003770480002 [PubMed: 14633332]
- Nesselroade, JR.; Molenaar, PCM. Pooling lagged covariance structures based on short, multivariate time-series for dynamic factor analysis. In: Hoyle, RH., editor. *Statistical strategies for small sample research*. Newbury Park, CA: Sage Publications; 1999. p. 223-250.
- Nielson F. The Flemish movement in Belgium after World War ii: a dynamic analysis. *American Journal of Economics and Sociology*. 1980; 45(1):76–94.
- Oravec Z, Tuerlinckx F, Vandekerckhove J. A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*. 2011; 16:468–490. [PubMed: 21823796]
- Oud, JHL. Comparison of four procedures to estimate the damped linear differential oscillator for panel data. In: Oud, J.; Satorra, A., editors. *Longitudinal models in the behavioral and related sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 2007.
- Oud JHL, Jansen RARG. Continuous time state space modeling of panel data by means of SEM. *Psychometrika*. 2000; 65(2):199–215.

- Oud JHL, Singer H. Special issue: Continuous time modeling of panel data. *Statistica neerlandica*. 2008; 62(1)doi: 10.1111/j.1467-9574.2007.00376.x
- Petzold LR. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM Journal of Scientific and Statistical Computing*. 1983; 4:136–148. DOI: 10.1137/0904010
- Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D. R Core Team. nlme: linear and nonlinear mixed effects models. 2014. R package version 3.1–118
- Ram, N.; Grimm, K. Growth curve modeling and longitudinal factor analysis. In: Overton, WF.; Molenaar, PCM., editors. *Handbook of child psychology and developmental science*. 7. Hoboken, NJ: Wiley; 2015. p. 758-788.
- Ram, N.; Shiyko, M.; Lunkenheimer, ES.; Doerksen, S.; Conroy, D. Families as coordinated symbiotic systems: Making use of nonlinear dynamic models. In A. In: Booth; McHale, S.; Landale, N., editors. *Emerging methods in family research, 4th national symposium on family issues*. Cham, Switzerland: Springer International Publishing; 2014. p. 19-37.
- Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach (with discussion). *Journal of Royal Statistical Society: Series B*. 2007; 69(5):741–796. DOI: 10.1111/j.1467-9868.2007.00610.x
- Ramsay, JO.; Hooker, G.; Graves, S. *Functional data analysis with R and MATLAB*. Springer-Verlag; New York: 2009.
- Ramsay, JO.; Silverman, BW. *Functional data analysis*. 2. Springer-Verlag; New York: 2005.
- Rosenfeld R. Race and sex differences in career dynamics. *American Sociological Review*. 1980; 45:583–609.
- Rothbart, MK.; Bates, JE. Temperament. In: Eisenberg, N.; Damon, W.; Lerner, RM., editors. *Handbook of child psychology: vol. 3, social, emotional, and personality development*. 6. Hoboken, NJ: John Wiley & Sons Inc; 2006. p. 99-166.
- Särkkä, S. *Bayesian filtering and smoothing*. Hillsdale, NJ: Cambridge University Press; 2013.
- SAS Institute Inc. *SAS 9.2 Help and Documentation*. Cary, NC: SAS Institute Inc; 2008.
- Shiyko MP, Lanza S, Tan X, Li R, Shiffman S. Using the time-varying effect model (TVEM) to examine dynamic associations between negative affect and self confidence on smoking urges: differences between successful quitters and relapsers. *Prevention Science*. 2012; 13(3):288–299. DOI: 10.1007/s11121-011-0264-z [PubMed: 22246429]
- Singer H. Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood vs. extended Kaman filter and itô-Taylor expansion. *Journal of Computational and Graphical Statistics*. 2002; 11:972–995. DOI: 10.1198/106186002808
- Singer H. Nonlinear continuous discrete filtering using kernel density estimates and functional integrals. *Journal of Mathematical Sociology*. 2003; 27:1–28. DOI: 10.1080/00222500305888
- Singer H. Sem modeling with singular moment matrices part i: ML-estimation of time series. *The Journal of Mathematical Sociology*. 2010; 34(4):301–320. DOI: 10.1080/0022250X.2010.532259
- Singer H. Sem modeling with singular moment matrices part ii: ml-estimation of sampled stochastic differential equations. *The Journal of Mathematical Sociology*. 2012; 36(1):22–43. DOI: 10.1080/0022250X.2010.532259
- Strogatz, SH. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Cambridge, MA: Westview; 1994.
- Tan X, Shiyko MP, Li R, Li Y, Dierker L. A time-varying effect model for intensive longitudinal data. *Psychological Methods*. 2012; 17(1):61–77. DOI: 10.1037/a0025814 [PubMed: 22103434]
- Thatcher, RW. A predator-prey model of human cerebral development. In: Newell, KM.; Molenaar, PCM., editors. *Applications of nonlinear dynamics to developmental process modeling*. Mahwah, NJ: Lawrence Erlbaum; 1998. p. 87-128.
- Thomas EA, Martin JA. Analyses of parent-infant interaction. *Psychological Review*. 1976; 83(2):141–156. DOI: 10.1037/0033-295X.83.2.141
- Tornøe C, Agersø H, Jonsson E, Madsen H, Nielsen H. Non-linear mixed-effects pharmacokinetic/ pharmacodynamic modelling in NLME using differential equations. *Computational Methods and Programs in Biomedicine*. 2004; 76(1):31–40. DOI: 10.1016/j.cmpb.2004.01.001

- Trail JB, Collins LM, Rivera DE, Li R, Piper ME, Baker TB. Functional data analysis for dynamical system identification of behavioral processes. *Psychological Methods*. 2013; 19:175–182. DOI: 10.1037/a0034035 [PubMed: 24079929]
- Tuma, NB.; Hannan, MT. *Social dynamics: Models and methods*. New York: Academic Press; 1984.
- van der Maas HLJ, Kolstein R, van der Pligt J. Sudden transitions in attitudes. *Sociological Methods & Research*. 2003; 32:125–152. DOI: 10.1177/0049124103253773
- Vaughn BE, Kopp CB, Krakow JB. The emergence and consolidation of self-control from eighteen to thirty months of age: Normative trends and individual differences. *Child Development*. 1984; 55:990–1004. DOI: 10.1037/0012-1649.22.6.752 [PubMed: 6734332]
- Voelkle MC, Oud JHL, Davidov E, Schmidt P. An sem approach to continuous time modeling of panel data: relating authoritarianism and anomia. *Psychological Methods*. 2012; 17:176–192. DOI: 10.1037/a0027543 [PubMed: 22486576]
- Volterra V. Variazioni e fluttuazioni del numero d'individui in specie animali con-viventi. *Mem Acad Lincei Roma*. 1926; 2:31–113. DOI: 10.1038/118558a0
- von Oertzen T, Boker SM. Time delay embedding increases estimation precision of models of intraindividual variability. *Psychometrika*. 2013; 75(1):158–175. DOI: 10.1007/s11336-009-9137-9 [PubMed: 23335820]
- Wood SN. Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*. 2003; 65(1):95–114. DOI: 10.1111/1467-9868.00374
- Zhu H, Ibrahim JG, Lee SY, Zhang HP. Appropriate perturbation and influence measures in local influence. *Annals of Statistics*. 2007; 35(6):2565–2588. DOI: 10.1214/009053607000000343
- Zill, DG. *A first course in differential equations*. 5. Boston: PWS-KENT Publishing Company; 1993.

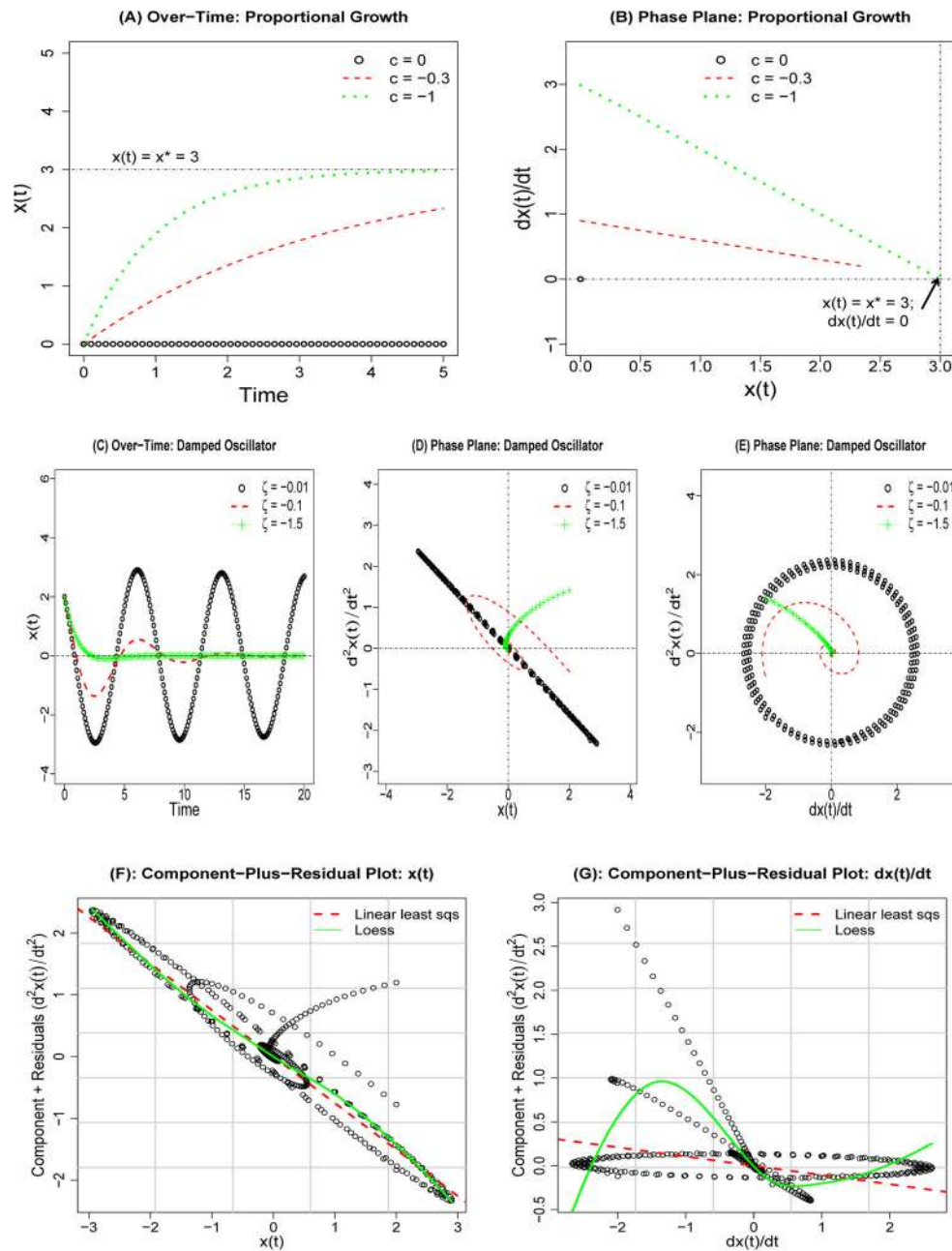


Figure 1. (A) Over-time trajectories of $x(t)$ from the proportional change model generated using based on a numerical ODE solver; (B) phase plane of levels and first derivatives in the proportional change model. The solid arrow points to one of the equilibrium points of the system at $x(t) = x^* = 3$, defined as points at which $\frac{dx}{dt} = 0$. (C) Over-time trajectories of $x(t)$ from the damped oscillator model generated using based on a numerical ODE solver at different values of ζ and η set to -0.8 ; (D) phase plane of levels and second derivatives in the damped oscillator model. (E) phase plane of first and second derivatives in the damped oscillator model (F)–

(G): component + residuals plot revealing the associations between the residuals and $x(t)$ and $dx(t)/dt$, respectively, with the linear effect of the other predictor partialled out.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

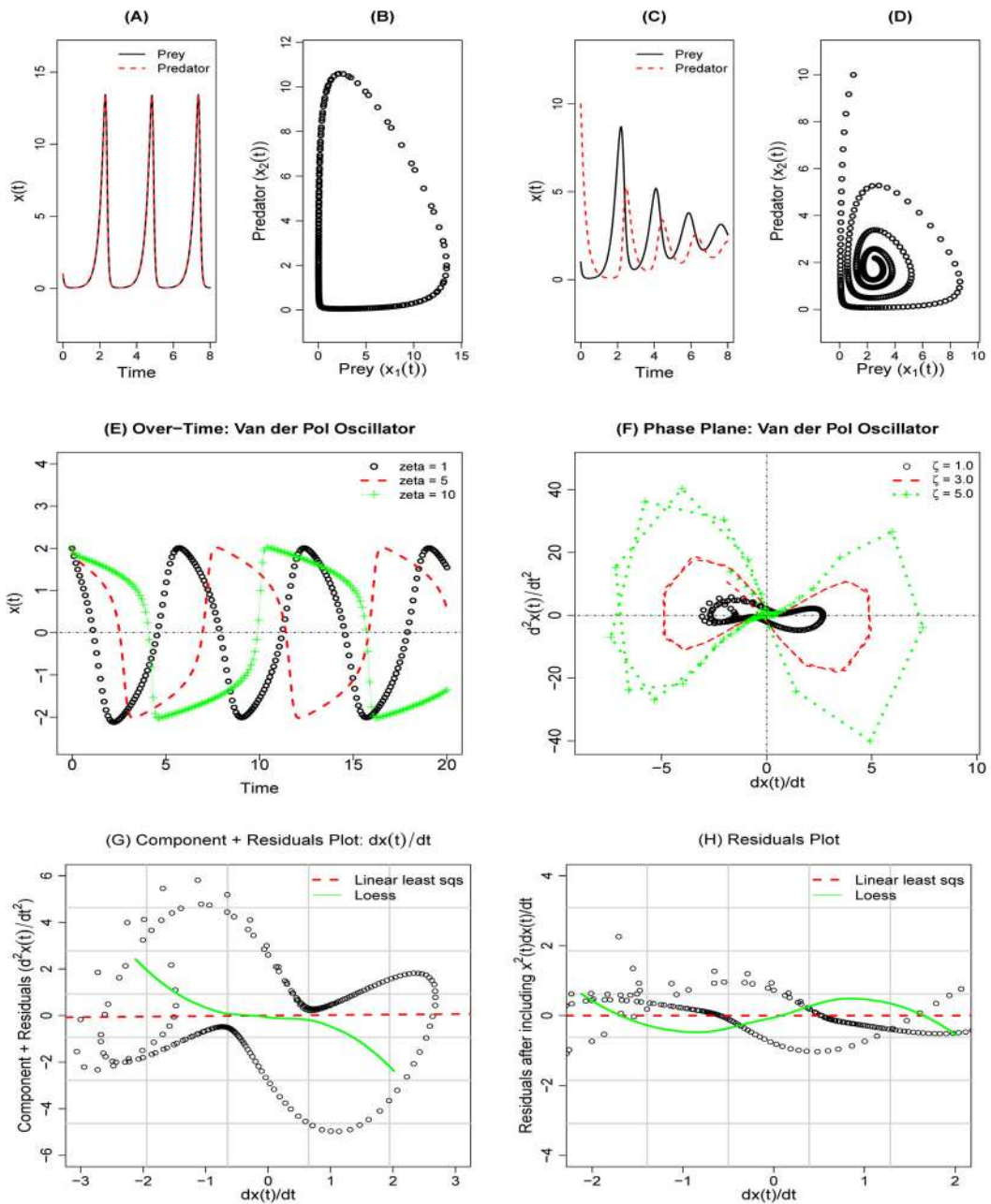


Figure 2.

(A)–(B) Over-time trajectories of the predator and prey population sizes from the classical predator-prey model and the corresponding phase plane plot generated with $r_1 = 4$, $r_2 = -3.5$, $a_{12} = -2$ and $a_{21} = 1.5$; (C)–(D) over-time trajectories of the two species’ sizes and the corresponding phase plane plot after adding a quadratic element, $-0.20x_1^2(t)$ to the prey equation in Equation 10; (E) Over-time trajectories of $x_1(t)$ from the van der Pol model generated with $\eta = 1$, $\zeta = 1.0, 3.0$, and 5.0 ; (F) the phase plane of the second and first derivatives; (G) component + residuals plot revealing the association between $d^2x(t)/dt^2$ and $dx(t)/dt$ after the linear effect of $x(t)$ has been partialled out; (H) plot of the residuals from an

expanded model (a model in which $x(t)$, $dx(t)/dt$ and $x^2(t)dx(t)/dt$ were included as predictors) against $dx(t)/dt$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

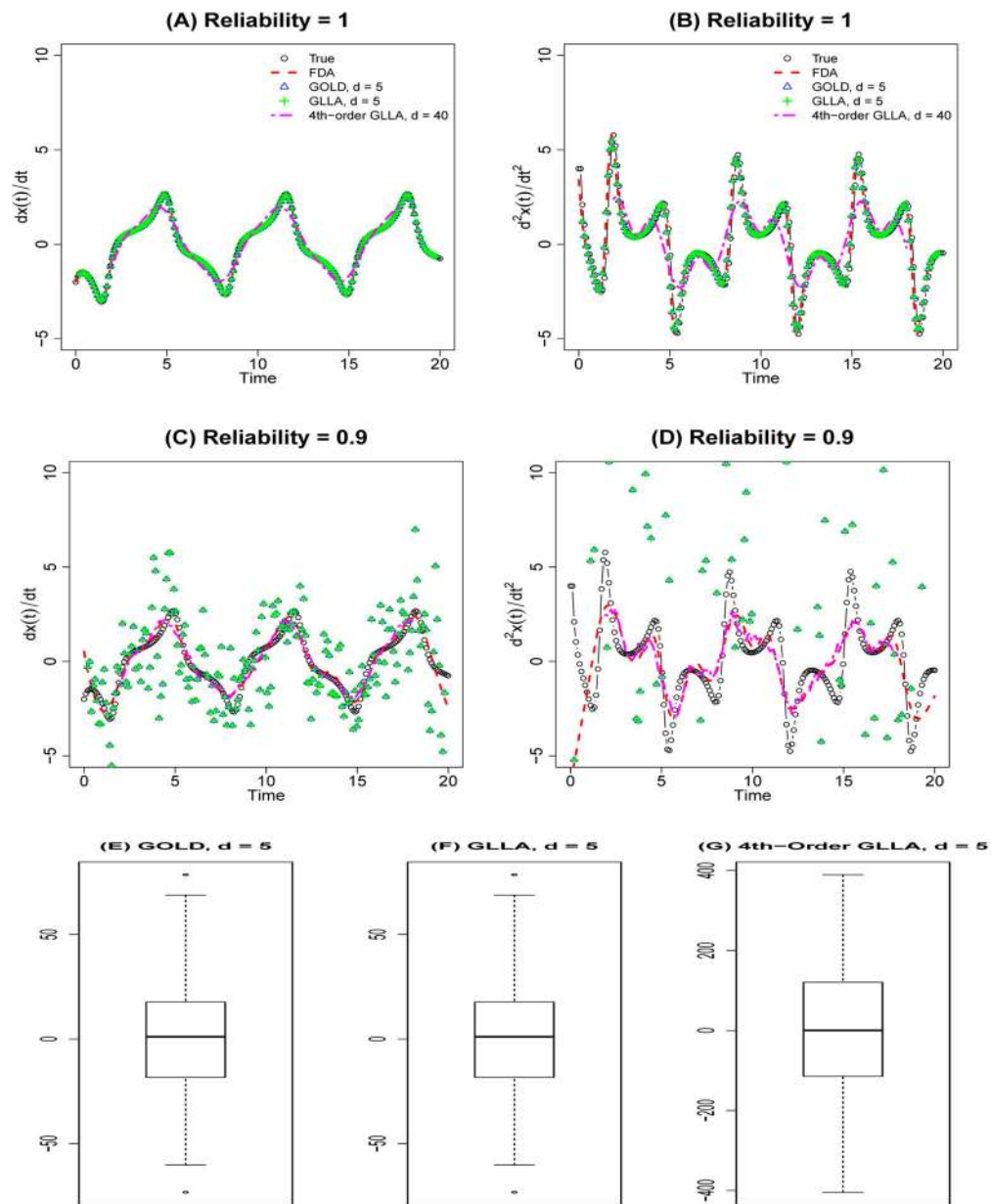


Figure 3.

(A) True and estimated first derivatives from the van der Pol oscillator model with no measurement noise; (B) true and estimated second derivatives with no measurement noise; (C)–(D) true and estimated first and second derivatives when reliability = 0.9. The lowest panels show boxplots of the second derivative estimates from the GOLD, the GLLA and the 4th-order GLLA when d , the embedding dimension, was set to 5. Here, reliability was computed as the ratio between the variance of the true $x(t)$ generated by means of a numerical ODE solver and the sum of the true $x(t)$ and measurement error variance. The same legends were used in plots (A)–(D) but are only shown in plots (A)–(B) due to space limitations.

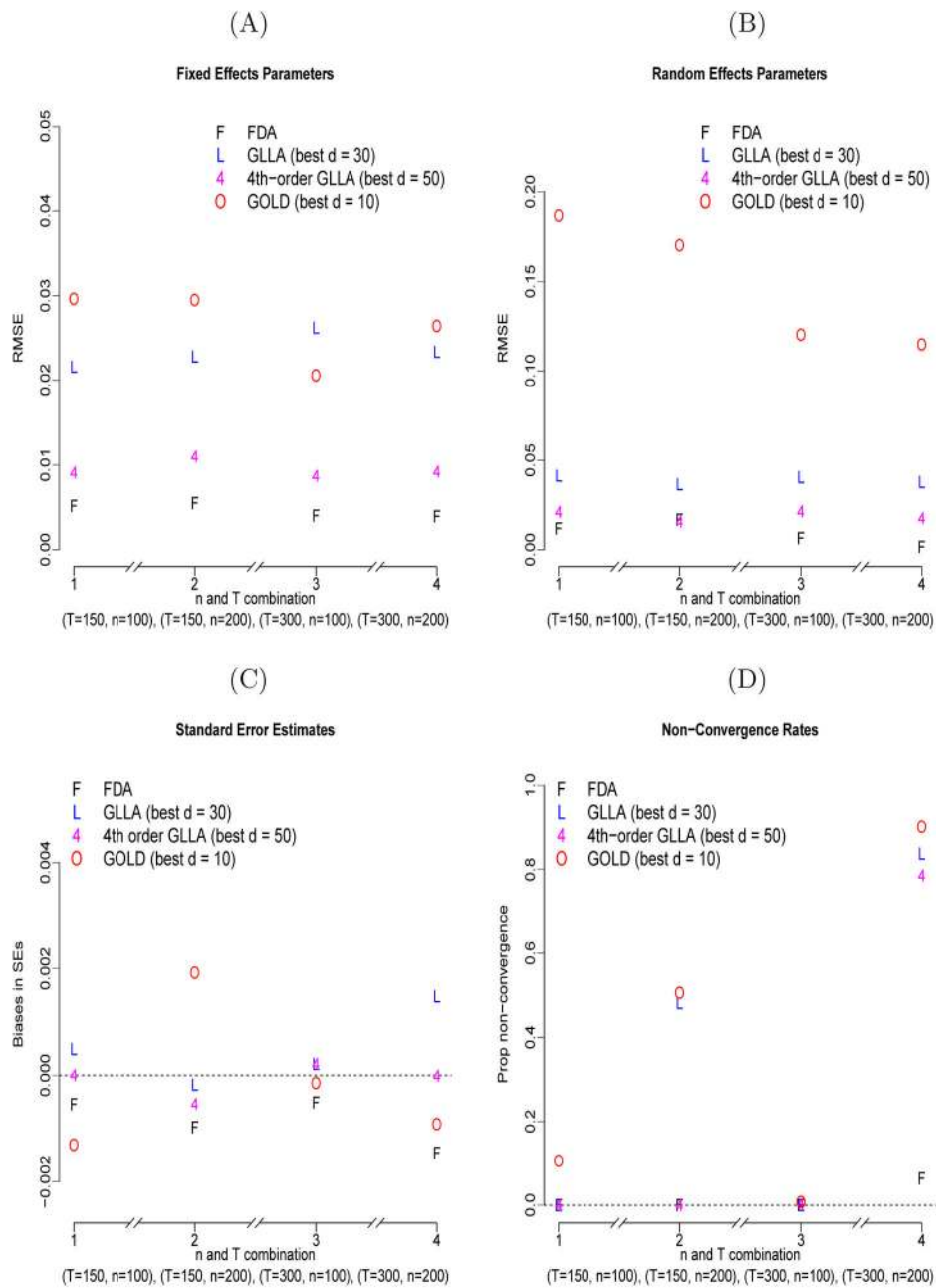


Figure 4. Results from fitting the coupled nonlinear oscillators model (see Equation 23) across sample size conditions, with (A) the RMSEs of the fixed effects point estimates; (B) the RMSEs of the random effect SD parameters, (C) biases in the SE estimates and (D) proportions of non-convergent replications.

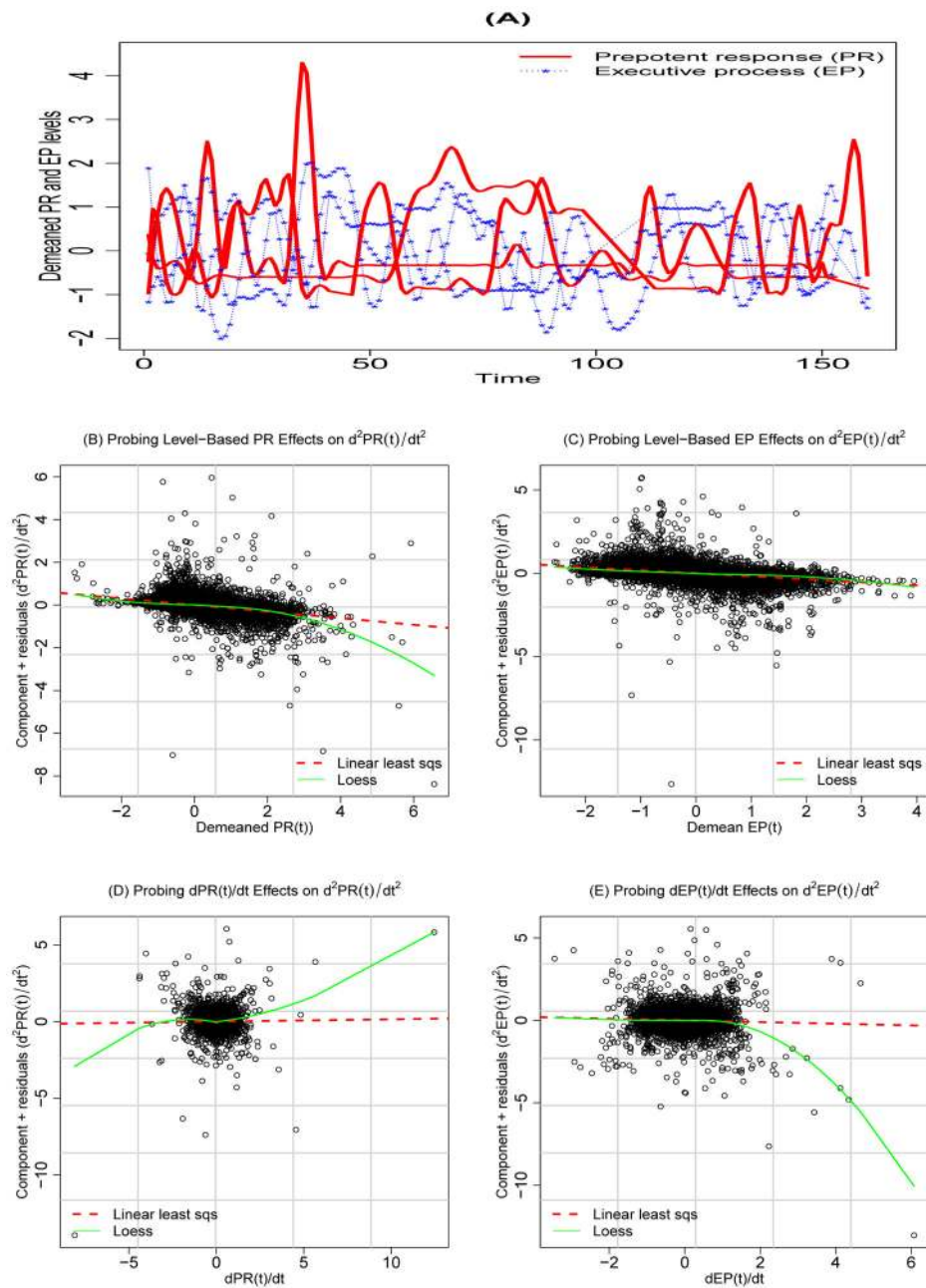


Figure 5. (A) Over-time trajectories of EP and PR from 5 randomly selected children; component + residuals plot revealing the partial associations between $d^2PR_i(t)/dt^2$ and: (B) demeaned $PR_i(t)$, (C) demeaned $EP_i(t)$, and (D) $\frac{dPR_i(t)}{dt}$; (E) the component + residuals plot revealing the partial associations between $\frac{d^2EP_i(t)}{dt^2}$ and $\frac{dEP_i(t)}{dt}$.

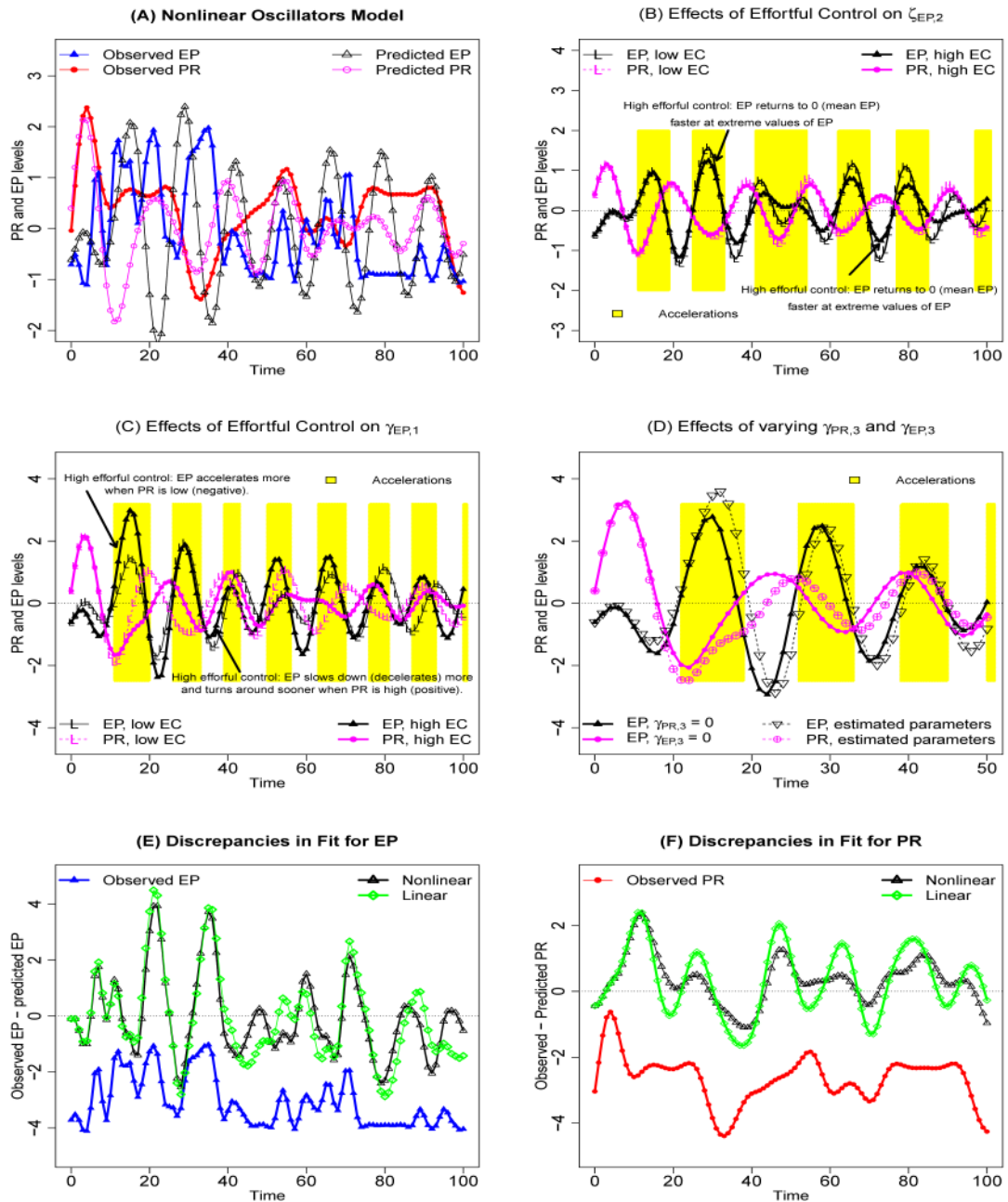


Figure 6.

(A) Observed data and predicted trajectories for one randomly selected participant generated using parameter estimates from the final model; (B)–(C) simulated trajectories illustrating the effects of $\gamma_{EP,1}$ and $\zeta_{EP,2}$, respectively, for a hypothetical child with high (+2SD above the sample average) and low (–2SD below the sample average) effortful control. (D) simulated trajectories generated using the estimated parameters from the final nonlinear model in comparison to trajectories generated by setting the values of $\gamma_{EP,3}$ and $\gamma_{PR,3}$ to zero; (E–F) discrepancies between the observed data and predicted trajectories for the same participant using the final nonlinear model and the linear comparison model, for EP and PR,

respectively. The shaded regions of plots (B)–(D) correspond to regions where the predicted EP trajectories marked with solid triangles show accelerations (positive second derivatives). The unshaded regions correspond to regions with decelerations (negative second derivatives) or second derivative value of zero (points of inflection). EC = effortful control.

Summary Statistics of Parameter Estimates for the Nonlinear Coupled Oscillators Model Using the FODEMixed with T = 150 and n = 100 across 500 MC Replications.

Table 1

True θ	Mean $\hat{\theta}$	RMSE	RB	SD	$a\widehat{SE}$	RDSE	Power/Type I error	Mean 95% CI
$\log(\eta_{10})$	0.41	0.26	-0.36	0.095	0.108	0.13	0.68	[0.05, 0.47]
η_{11}	-0.50	-0.44	-0.12	0.037	0.042	0.15	1.00	[-0.52, -0.36]
η_{12}	0.00	-0.01	—	0.084	0.097	0.16	0.01	[-0.20, 0.18]
$\log(\eta_{20})$	0.00	-0.08	—	0.076	0.099	0.31	0.93	[-0.27, 0.12]
η_{21}	0.00	-0.00	—	0.030	0.039	0.30	0.00	[-0.08, 0.07]
η_{22}	-0.30	-0.24	-0.20	0.072	0.089	0.22	0.78	[-0.41, -0.07]
a_{12}	-0.02	-0.01	-0.35	0.004	0.001	-0.83	1.00	[-0.01, -0.01]
a_{21}	-0.03	-0.02	-0.39	0.004	0.001	-0.83	1.00	[-0.02, -0.02]
σ_{η_1}	0.50	0.46	-0.08	0.034	—	—	1.00	[0.40, 0.54]
σ_{η_2}	0.50	0.43	-0.14	0.027	—	—	1.00	[0.37, 0.50]
σ_{ε_1}	1.00	1.23 ^a	0.23	0.126	—	—	1.00	[1.22, 1.25]
σ_{ε_2}	1.00	1.52 ^a	0.52	0.130	—	—	1.00	[1.48, 1.57]

^aThese estimates correspond to the residual SD parameters $\sigma_{\varepsilon_1}^*$ and $\sigma_{\varepsilon_2}^*$ in \sum_{ε}^* depicted in Equation 16. True θ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{G} \sum_{g=1}^G \hat{\theta}_g$, where $\hat{\theta}_g$ = estimate of θ from the g th MC replication; $RMSE = \sqrt{\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - true \theta)^2}$; RB=relative bias= $\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - true \theta) / true \theta$; SE = standard deviation of $\hat{\theta}$ across MC runs;

$a\widehat{SE}$ =average standard error estimate across MC 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 MC replications; Mean 95% CI = 95% CI constructed from the average lower and upper 95% limits reported by *nIme* obtained across MC replications.

Summary Statistics of Parameter Estimates for the Nonlinear Coupled Oscillators Model Using the FODEMixed with T = 150 and n = 200 across 500 MC Replications.

Table 2

True θ	Mean $\hat{\theta}$	RMSE	RB	SD	$a\widehat{SE}$	RDSE	Power/Type I error	Mean 95% CI
$\log(\eta_0)$	0.41	0.24	-0.40	0.059	0.077	0.30	0.92	[0.09, 0.39]
η_1	-0.50	-0.44	-0.13	0.022	0.030	0.34	1.00	[-0.50, -0.38]
η_2	0.00	-0.00	—	0.056	0.068	0.23	0.00	[-0.14, 0.13]
$\log(\eta_{20})$	0.00	-0.09	—	0.056	0.071	0.26	0.82	[-0.22, 0.05]
η_{21}	0.00	-0.00	—	0.022	0.027	0.24	0.00	[-0.06, 0.05]
η_{22}	-0.30	-0.23	-0.22	0.046	0.063	0.36	1.00	[-0.36, -0.11]
a_{12}	-0.02	-0.01	-0.34	0.003	0.001	-0.84	1.00	[-0.01, -0.01]
a_{21}	-0.03	-0.02	-0.39	0.003	0.000	-0.84	1.00	[-0.02, -0.02]
σ_{η_1}	0.50	0.47	-0.06	0.025	—	—	1.00	[0.42, 0.52]
σ_{η_2}	0.50	0.44	-0.13	0.020	—	—	1.00	[0.39, 0.48]
σ_{ε_1}	1.00	1.22 ^a	0.22	0.084	—	—	1.00	[1.21, 1.23]
σ_{ε_2}	1.00	1.53 ^a	0.53	0.092	—	—	1.00	[1.50, 1.56]

^aThese estimates correspond to the residual SD parameters $\sigma_{\varepsilon_1}^*$ and $\sigma_{\varepsilon_2}^*$ in \sum_{ε}^* depicted in Equation 16. True θ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{G} \sum_{g=1}^G \hat{\theta}_g$, where $\hat{\theta}_g$ = estimate of θ from the g th MC replication; $RMSE = \sqrt{\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - true \theta)^2}$; RB = relative bias = $\frac{1}{H} \sum_g (\hat{\theta}_g - true \theta) / true \theta$; SE = standard deviation of $\hat{\theta}$ across MC runs;

$a\widehat{SE}$ = average standard error estimate across MC 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 MC replications; Mean 95% CI = 95% CI constructed from the average lower and upper 95% limits reported by *nIme* obtained across MC replications.

Summary Statistics of Parameter Estimates for the Nonlinear Coupled Oscillators Model Using the FODEMixed with T = 300 and n = 100 across 500 MC Replications.

Table 3

	True θ	Mean $\hat{\theta}$	RMSE	RB	SD	$a\widehat{SE}$	RDSE	Power/Type I error	Mean 95% CI	
	$\log(\eta_0)$	0.41	0.33	0.07	-0.18	0.087	0.107	0.24	0.89	[0.12, 0.54]
	η_1	-0.50	-0.47	0.03	-0.06	0.034	0.042	0.23	1.00	[-0.55, -0.39]
	η_2	0.00	-0.00	0.00	—	0.079	0.096	0.22	0.00	[-0.19, 0.19]
	$\log(\eta_0)$	0.00	-0.03	0.03	—	0.082	0.105	0.29	1.00	[-0.23, 0.18]
	η_1	0.00	-0.01	0.01	—	0.033	0.041	0.26	0.00	[-0.09, 0.07]
	η_2	-0.30	-0.27	0.03	-0.11	0.071	0.094	0.32	0.85	[-0.45, -0.08]
	a_{12}	-0.02	-0.01	0.01	-0.30	0.004	0.000	-0.88	1.00	[-0.01, -0.01]
	a_{21}	-0.03	-0.02	0.01	-0.30	0.003	0.000	-0.88	1.00	[-0.02, -0.02]
	σ_{η_1}	0.50	0.47	0.03	-0.06	0.028	—	—	1.00	[0.41, 0.54]
	σ_{η_2}	0.50	0.46	0.04	-0.09	0.025	—	—	1.00	[0.40, 0.52]
	σ_{ε_1}	1.00	1.09 ^a	0.09	0.09	0.134	—	—	1.00	[1.08, 1.09]
	σ_{ε_2}	1.00	1.33 ^a	0.33	0.33	0.134	—	—	1.00	[1.30, 1.35]

^aThese estimates correspond to the residual SD parameters $\sigma_{\varepsilon_1}^*$ and $\sigma_{\varepsilon_2}^*$ in \sum_{ε}^* depicted in Equation 16. True (θ) true value of a parameter; Mean $\hat{\theta} = \frac{1}{G} \sum_{g=1}^G \hat{\theta}_g$, where $\hat{\theta}_g$ = estimate of θ from the g th MC replication; $\text{RMSE} = \sqrt{\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - \text{true } \theta)^2}$; RB = relative bias = $\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - \text{true } \theta) / \text{true } \theta$; SE = standard deviation of θ across MC runs;

$a\widehat{SE}$ = average standard error estimate across MC 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 MC replications; Mean 95% CI = 95% CI constructed from the average lower and upper 95% limits reported by *nIme* obtained across MC replications.

Summary Statistics of Parameter Estimates for the Nonlinear Coupled Oscillators Model Using the FODEMixed with T = 300 and n = 200 across 500 MC Replications.

Table 4

	True θ	Mean $\hat{\theta}$	RMSE	RB	SD	$a\widehat{SE}$	RDSE	Power/ Type I error	Mean 95% CI
$\log(\eta_0)$	0.41	0.33	0.08	-0.20	0.053	0.077	0.46	1.00	[0.17, 0.48]
η_1	-0.50	-0.47	0.03	-0.07	0.023	0.030	0.33	1.00	[-0.52, -0.41]
η_2	0.00	-0.00	0.00	—	0.054	0.069	0.29	0.00	[-0.14, 0.13]
$\log(\eta_0)$	0.00	-0.04	0.04	—	0.060	0.073	0.23	0.99	[-0.18, 0.11]
η_1	0.00	-0.00	0.00	—	0.023	0.028	0.24	0.00	[-0.06, 0.05]
η_2	-0.30	-0.27	0.03	-0.11	0.051	0.066	0.28	1.00	[-0.39, -0.14]
θ_{12}	-0.02	-0.01	0.01	-0.27	0.003	0.000	-0.88	1.00	[-0.02, -0.01]
θ_{21}	-0.03	-0.02	0.01	-0.29	0.002	0.000	-0.88	1.00	[-0.02, -0.02]
σ_{η_1}	0.50	0.48	0.02	-0.04	0.021	—	—	1.00	[0.43, 0.53]
σ_{η_2}	0.50	0.46	0.04	-0.09	0.019	—	—	1.00	[0.41, 0.51]
σ_{ε_1}	1.00	1.09 ^a	0.09	0.09	0.094	—	—	1.00	[1.09, 1.10]
σ_{ε_2}	1.00	1.33 ^a	0.33	0.33	0.097	—	—	1.00	[1.31, 1.35]

^aThese estimates correspond to the residual SD parameters $\sigma_{\varepsilon_1}^*$ and $\sigma_{\varepsilon_2}^*$ in \sum_{ε}^* depicted in Equation 16. True θ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{G} \sum_{g=1}^G \hat{\theta}_g$ where $\hat{\theta}_g$ = estimate of θ from the g th MC replication; $\text{RMSE} = \sqrt{\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - \text{true } \theta)^2}$; RB=relative bias = $\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - \text{true } \theta) / \text{true } \theta$; SE = standard deviation of $\hat{\theta}$ across MC runs;

$a\widehat{SE}$ =average standard error estimate across MC 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 MC replications; Mean 95% CI = 95% CI constructed from the average lower and upper 95% limits reported by *nIme* obtained across MC replications.

Table 5
 Summary Statistics of Parameter Estimates using the FODEmixed and GLLA across 500 MC Replications.

d	RMSE fixed	RMSE rand	RB fixed	RB rand	SE biases	Power	Non-cony
<u>T=150, n=100</u>							
FODEmixed	0.01	0.01	-0.04	0.02	-0.00	0.97	0.00
GLLA							
5.00	0.53	0.30	0.89	0.60	0.53	0.68	0.42
10.00	0.17	0.12	0.21	0.24	0.17	0.90	0.02
30.00	0.02	0.04	-0.10	-0.08	-0.00	0.97	0.00
40.00	0.05	0.06	-0.17	-0.12	-0.02	0.95	0.00
50.00	0.07	0.08	-0.24	-0.16	-0.04	0.91	0.00
<u>T=150, n=200</u>							
FODEmixed	0.01	0.02	-0.05	0.03	-0.00	1.00	0.00
GLLA							
5.00	0.53	0.32	0.88	0.64	0.53	0.74	0.74
10.00	0.17	0.13	0.20	0.27	0.17	0.96	0.49
30.00	0.02	0.04	-0.10	-0.07	-0.00	1.00	0.48
40.00	0.05	0.06	-0.17	-0.11	-0.02	0.99	0.49
50.00	0.07	0.08	-0.25	-0.15	-0.04	0.95	0.00
<u>T=300, n=100</u>							
FODEmixed	0.00	0.01	-0.05	-0.01	-0.00	0.97	0.00
GLLA							
5.00	0.52	0.32	0.87	0.64	0.52	0.71	0.65
10.00	0.16	0.14	0.18	0.29	0.16	0.93	0.15
30.00	0.03	0.04	-0.11	-0.08	-0.01	0.96	0.00
40.00	0.05	0.06	-0.18	-0.12	-0.02	0.94	0.00
50.00	0.08	0.08	-0.25	-0.16	-0.04	0.91	0.00
<u>T=300, n=200</u>							
FODEmixed	0.00	0.00	-0.04	-0.00	-0.00	1.00	0.06
GLLA							

d	RMSE fixed	RMSE rand	RB fixed	RB rand	SE biases	Power	Non-conv
5.00	0.52	0.34	0.86	0.67	0.52	0.81	0.72
10.00	0.16	0.14	0.19	0.28	0.16	0.97	0.90
30.00	0.02	0.04	-0.10	-0.08	-0.00	1.00	0.84
40.00	0.04	0.06	-0.16	-0.12	-0.02	0.99	0.85
50.00	0.07	0.08	-0.25	-0.15	-0.04	0.95	0.85

RMSE fixed and RMSE rand = RMSE as averaged across all fixed effects and all random effects SD parameters, respectively; $RMSE = \sqrt{\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - true \theta)^2}$; RB fixed and RB rand = relative bias as averaged across all fixed effects and all random effects SD parameters, respectively; $RB = \text{relative bias} = \frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - true \theta) / true \theta$; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the MC replications as averaged across all parameters whose true population values were not zero; Non-conv = proportion of non-convergent cases. The FODEmixed estimates, the best GLLA estimates plotted in Figures 4, and the associated embedding dimension are highlighted in bold.

Table 6
 Summary Statistics of Parameter Estimates using the Fourth-Order GLLA across 500 MC Replications.

d	RMSE fixed	RMSE rand	RB fixed	RB rand	SE biases	Power	Non-cony
<u>T = 150, n = 100</u>							
FODEmixed	0.01	0.01	-0.04	0.02	-0.00	0.97	0.00
<u>4th order GLLA</u>							
40.00	0.02	0.01	-0.00	-0.03	0.02	0.98	0.00
50.00	0.01	0.02	-0.03	-0.04	0.01	0.98	0.00
<u>T = 150, n = 200</u>							
FODEmixed	0.01	0.02	-0.05	0.03	-0.00	1.00	0.00
<u>4th-order GLLA</u>							
40.00	0.02	0.01	-0.00	-0.02	0.02	1.00	0.49
50.00	0.01	0.02	-0.04	-0.03	0.01	1.00	0.00
<u>T = 300, n = 100</u>							
FODEmixed	0.00	0.01	-0.05	-0.01	-0.00	0.97	0.00
<u>4th-order GLLA</u>							
40.00	0.02	0.01	-0.02	-0.02	0.02	0.98	0.00
50.00	0.01	0.02	-0.04	-0.04	0.01	0.98	0.00
<u>T = 300, n = 200</u>							
FODEmixed	0.00	0.01	-0.05	-0.01	-0.00	0.97	0.00
<u>4th order GLLA</u>							
40.00	0.02	0.01	-0.01	-0.01	0.02	1.00	0.79
50.00	0.01	0.02	-0.04	-0.04	0.01	1.00	0.79

RMSE fixed and RMSE rand = RMSE as averaged across all fixed effects and all random effects SD parameters, respectively; $RMSE = \sqrt{\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - true \theta)^2}$; RB fixed and RB random = relative bias as averaged across all fixed effects and all random effects SD parameters, respectively; $RB = \text{relative bias} = \frac{\sum_{g=1}^G (\hat{\theta}_g - true \theta)}{\sum_{g=1}^G true \theta}$; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the MC replications as averaged across all parameters whose true population values were not zero; Non-cony = proportion of non-convergent cases. The FODEmixed estimates, the best fourth-order GLLA estimates plotted in Figures 4, and the associated embedding dimension are highlighted in bold.

Table 7

Summary Statistics of Parameter Estimates using the GOLD across 500 MC Replications.

d	RMSE fixed	RMSE rand	RB fixed	RB rand	SE biases	Power	Non-cony
<u>T = 150, n = 100</u>							
FODEmixed	0.01	0.01	-0.04	0.02	-0.00	0.97	0.00
GOLD							
5.00	0.02	0.50	-0.19	-1.00	0.27	0.46	0.79
10.00	0.03	0.19	0.03	-0.37	-0.00	0.89	0.11
30.00	0.04	0.04	0.19	0.07	-0.00	0.98	0.00
40.00	0.08	0.10	0.38	0.20	-0.02	0.98	0.01
50.00	0.06	1.71	0.29	3.41	-0.80	0.95	0.20
<u>T = 150, n = 200</u>							
FODEmixed	0.01	0.02	-0.05	0.03	-0.00	1.00	0.00
GOLD							
5.00	0.03	0.50	-0.20	-1.00	0.19	0.54	0.91
10.00	0.03	0.17	0.02	-0.34	0.00	0.99	0.51
30.00	0.04	0.04	0.19	0.09	-0.00	1.00	0.48
40.00	0.08	0.15	0.37	0.31	-0.01	1.00	0.50
50.00	0.06	1.97	0.37	3.95	-0.23	0.97	0.23
<u>T = 300, n = 100</u>							
FODEmixed	0.00	0.01	-0.05	-0.01	-0.00	0.97	0.00
GOLD							
5.00	0.03	0.50	-0.16	-0.99	0.15	0.55	0.87
10.00	0.02	0.12	-0.01	-0.24	-0.00	0.96	0.01
30.00	0.04	0.04	0.18	0.07	-0.00	0.98	0.00
40.00	0.07	0.38	0.30	0.75	-0.32	0.98	0.03
50.00	0.71	12.09	-2.09	24.18	-10.10	0.95	0.32
<u>T = 300, n = 200</u>							
FODEmixed	0.00	0.01	-0.05	-0.01	-0.00	0.97	0.00
GOLD							

d	RMSE fixed	RMSE rand	RB fixed	RB rand	SE biases	Power	Non-conv
5.00	0.02	0.50	-0.18	-0.99	0.11	0.58	0.83
10.00	0.03	0.11	-0.00	-0.23	-0.00	1.00	0.90
30.00	0.04	0.04	0.18	0.08	-0.00	1.00	0.34
40.00	0.07	0.45	0.30	0.90	-0.08	0.99	0.87
50.00	0.48	18.65	-0.64	37.30	-1.05	0.97	0.90

RMSE fixed and RMSE rand = RMSE as averaged across all fixed effects and all random effects SD parameters, respectively; $RMSE = \sqrt{\frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - true \theta)^2}$; RB fixed and RB rand = relative bias as averaged across all fixed effects and all random effects SD parameters, respectively; $RB = \text{relative bias} = \frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - true \theta) / true \theta$; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the MC replications as averaged across all parameters whose true population values were not zero; Non-conv = proportion of non-convergent cases. The FODEmixed estimates, the best GOLD estimates plotted in Figures 4, and the associated embedding dimension are highlighted in bold.

Table 8

Parameter Estimates from the Empirical Illustrative Example.

Parameters	Estimates (SE)	95% CI
Intercept of $\eta_{EP,1,i}$	-0.23 (0.004) **	[-0.24, -0.22]
Intercept of $\eta_{PR,1,i}$	-0.16 (0.004) **	[-0.16, -0.15]
Intercept of $\zeta_{EP,1,i}$	-0.02 (0.01) ^(NS)	[-0.04, 0.01]
Intercept of $\zeta_{PR,1,i}$	-0.01 (0.02) ^(NS)	[-0.04, 0.03]
Intercept of $\zeta_{EP,2,i}$	0.01 (0.01) ^(NS)	[-0.01, 0.04]
Intercept of $\zeta_{PR,2,i}$	0.001 (0.01) ^(NS)	[-0.02, 0.02]
Effect of effortful control on $\zeta_{EP,2,i}$	-0.04(0.02) *	[-0.08, -0.01]
Intercept of $\gamma_{EP,1,i}$	-0.09(0.01) **	[-0.10, -0.08]
Effect of effortful control on $\gamma_{EP,1,i}$	-0.02(0.01) **	[-0.04, -0.01]
Intercept of $\gamma_{EP,3,i}$	-0.04(0.01) **	[-0.05, -0.03]
Intercept of $\gamma_{EP,4,i}$	-0.01(0.01) ^(NS)	[-0.04, 0.02]
Intercept of $\gamma_{PR,1,i}$	-0.03(0.01) **	[-0.04, -0.02]
Intercept of $\gamma_{PR,3,i}$	0.02(0.01) ^(*)	[0.002, 0.04]
Intercept of $\gamma_{PR,4,i}$	-0.04(0.02) ^(*)	[-0.07, -0.01]
Random effect SD for $\zeta_{EP,1,i}$	0.08(-) *	[0.06, 0.10]
Random effect SD for $\zeta_{EP,2,i}$	0.06(-) *	[0.03, 0.10]
Random effect SD for $\zeta_{PR,1,i}$	0.17(-) *	[0.14, 0.20]
Random effect SD for $\zeta_{PR,2,i}$	0.06(-) *	[0.05, 0.08]
Random effect SD for $\gamma_{EP,1,i}$	0.03(-) *	[0.03, 0.05]
Random effect SD for $\gamma_{EP,3,i}$	0.05(-) *	[0.04, 0.07]
Random effect SD for $\gamma_{EP,4,i}$	0.11(-) *	[0.09, 0.14]
Random effect SD for $\gamma_{PR,1,i}$	0.04(-) *	[0.03, 0.05]
Random effect SD for $\gamma_{PR,3,i}$	0.10(-) *	[0.09, 0.13]
Random effect SD for $\gamma_{PR,4,i}$	0.12(-) *	[0.10, 0.15]
$\sigma_{\varepsilon_{PR}}^*$	0.37(-) *	[0.37, 0.38]
$\sigma_{\varepsilon_{EP}}^*$	0.29(-) *	[0.28, 0.30]

Note.

* $p < .05$;

** $p < .01$.

(NS) Even though the some of the intercept terms were not statistically significant, they were retained in the final model in light of the substantial interindividual differences observed in these parameters.

$\sigma_{\varepsilon_{PR}}^*$ and $\sigma_{\varepsilon_{EP}}^*$ are residual SDs from fitting the nonlinear mixed effects model to the demeaned PR and EP scores. Other parameters are as defined in the context of the ODE model in Equation 24.