

Practical Tools and Guidelines for Exploring and Fitting Linear and Nonlinear Dynamical Systems Models

Sy-Miin Chow

The Pennsylvania State University

In press, *Multivariate Behavioral Research*

Abstract

A dynamical system is a system of variables that show some regularity in how they involve over time. Change concepts described in most dynamical systems models are by no means novel to social and behavioral scientists, but most applications of dynamic modeling techniques in these disciplines are grounded on a narrow subset of — typically linear — theories of change. I provide practical guidelines, recommendations, and software code for exploring and fitting dynamical systems models with linear and nonlinear change functions in the context of four illustrative examples. Cautionary notes, challenges, and unresolved issues in utilizing these techniques are discussed.

Keywords: Differential equation, functional data analysis, dynamical systems, GAM, time-varying parameters

A dynamical system is, broadly speaking, a system of variables that show some regularity in how they involve over time. This admittedly broad and somewhat abstract definition is in part a reflection of the changing definition of dynamical systems in the literature in the past few decades. For instance, Scheinerman (1996, p. 1) referred to a dynamical system as “a system that is doing the same thing repeatedly” and one that “always knows what it is going to do next.” One alternative definition, offered by Boker and Nesselroade (2002), describes dynamical systems as systems that change over time such that their current states are somehow dependent upon their previous states. This definition does not emphasize the notion of perfect predictability, but it does impose some constraints on the type of systems that may be considered as dynamical systems. Thus, a system that shows a constant amount of change (i.e., a linear true change trajectory) would be considered as a dynamical system under Scheinerman’s, but not Boker’s definition. As such, the definition

Funding for this study was provided in part by NSF grant SES-1357666, NIH grant R01GM105004 and by the Raymond B. Cattell Award from the Society of Multivariate Experimental Psychology. The author is indebted to Dr. Steve West, other anonymous reviewers, and various colleagues and students in the QuantDev group of Penn State for valuable comments on earlier drafts of this manuscript. Correspondence concerning this article can be addressed to Sy-Miin Chow, the Pennsylvania State University, 409 Biobehavioral Health Building, University Park, PA 16802 or by email to symiin@psu.edu.

of dynamical systems is itself “dynamic” because it varies in breadth and in some selected properties depending on the specific tools, and relatedly, modeling assumptions adopted by individual researchers.

Scheinerman’s definition targets a class of systems whose patterns of regularity may be extracted, elucidated, and predicted perfectly in the future given the right tools. Systems with perfect predictability are known as *deterministic systems*. That is, given perfect knowledge of their previous values and the rules that govern their changes, such systems’ future values can be predicted perfectly despite their seemingly complex observed patterns of change. The belief that such systems exist, and may have change functions that can be controlled experimentally, has fueled much of the enthusiasm in earlier research on dynamical systems. With more researchers collaborating across disciplines and contributing collectively to the development of dynamical systems analytic tools, more common grounds have since been discovered between dynamical systems tools and other well-known techniques, such as time series analysis, differential and difference equation modeling, and latent variable modeling. Consequently, more tools are now available for modeling *stochastic systems*, namely, systems that show regularity, but also some uncertainties in how they change over time. These tools have helped expand what researchers typically think of as dynamical systems. Regardless of whether a researcher endorses the notion of deterministic or stochastic change mechanisms, extraction and examination of the regularity in a system’s dynamics have remained one of the core features of dynamical systems-inspired techniques.

Dynamical systems modeling tools are unique in that they require researchers to identify the change mechanism that dictates *in what ways* the system has changed, and any regularity and heterogeneity therein. Earlier dynamical systems approaches tended to emphasize exploratory tools that help reveal the graphical relations among the variables that constitute a dynamical system (Kaplan & Glass, 1995). These tools have some unique strengths and natural appeal in scenarios involving nonlinear systems, especially those whose long-term behaviors are difficult or impossible to infer analytically. They are not always as powerful in discerning the dynamics embedded in the kinds of longitudinal data commonly encountered in the social and behavioral sciences — namely, data that are noisy, of finite time lengths, involve multiple replications across individuals (or other units of analysis), and are possibly heterogeneous within and/or across units over time.

My goal in this article is to provide practical demonstrations and recommendations on how standard graphical and inferential tools in the regression and related frameworks can be used to clarify the change mechanisms characterizing dynamical systems, using data that more closely mirror the kinds of longitudinal data commonly available in the social and behavioral sciences. In the remainder of this article, I will first introduce differential equations (DEs) as ways to characterize change mechanisms. This is followed by four didactic demonstrations of how tools for derivative estimation can be used in conjunction with regression and related modeling tools to explore and construct DE models. To provide benchmark comparisons to the results of exploratory techniques, one possible way of fitting confirmatory differential equation models is described. Practical guidelines, software code, recommendations, and cautionary notes in using these exploratory and model-fitting tools are discussed.

Differential Equation (DE) Models

One possible mechanism for modeling the recurrence in a system’s dynamics over time

is through DE modeling. DE modeling has been widely employed as a modeling tool in the physical sciences, engineering, and many other scientific disciplines long before the advent of dynamical systems analytic approaches. As a class, DE models provide a framework for expanding the quest for describing *whether* people have changed to *how* they have changed. In practice, DE 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; Hawkley, Burleson, Berntson, & Cacioppo, 2003; Merrilees, Goeke-Morey, & Cummings, 2008). In the social and behavioral sciences, researchers have also utilized DEs to capture the dynamic aspects of social processes, organizations, and human behaviors (e.g., Arminger, 1986; Coleman, 1968; Tuma & Hannan, 1984).

Given repeated observations of a continuous process, $y_i(t)$, for unit (e.g., person) i ($i = 1, \dots, n$) at times $t = 0, \dots, T$, the first derivative (rate of change) of y at any arbitrary time point is the change in y that occurs within a (infinitely) small window of time, a . Formally,

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

The notation “ (t) ” is a continuous index of time that may take on any real value. A positive value of derivative $\frac{dy_i(t)}{dt}$ reflects an instantaneous increase or growth in the value of $y_i(t)$ at that particular moment; a negative value of $\frac{dy_i(t)}{dt}$ indicates an instantaneous decline. If $\frac{dy_i(t)}{dt} = 0$, then $y_i(t)$ is said to be “static” or manifest no change at that particular moment. Generally, fixed points for any dynamical system are values of the dependent variables for which the rates of change for all its constituent dependent variables, including derivative variables, are all equal to zero.

In a similar vein, the second derivative is the change in the rate of change of y :

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

where this change in instantaneous growth or decline may be conceptualized as the curvature in $y_i(t)$. Higher positive values of $\frac{d^2 y_i(t)}{dt^2}$ signify greater or more precipitous growth or decline in the level of $y_i(t)$, much like acceleration occurs when a car travels successively greater distances at close intervals; negative values of $\frac{d^2 y_i(t)}{dt^2}$ indicate reductions in growth or decline, much like deceleration in a car as travels decreasing distances. Other higher-order derivatives then capture further changes in these change qualities. All of the variables that determine the current values of a system, including level and derivative variables, constitute the *state space* of the dynamical system. The order of a DE model, denoted as m below, indicates the highest-order derivative in a model.

When multiple dependent variables are involved or in systems that are of higher orders than 1 (i.e., involving higher-order derivatives beyond first derivatives), it is often convenient to gather the functions of derivatives into a vector form as:

$$\frac{d\mathbf{y}_i(t)}{dt} = \mathbf{f}[\mathbf{y}_i(t), \mathbf{u}_i(t)], \quad i = 1, \dots, n; t = 0, \dots, T. \quad (1)$$

In Equation (1), $\frac{d}{dt}(\cdot)$ is a differential operator that takes the derivative of the element enclosed in parentheses with respect to time, and $\mathbf{y}_i(t)$ is now a vector of variables of

interest containing level and derivatives at time t that are of lower order than m . $\mathbf{f}[\cdot]$ is a vector of functions, typically referred to as drift functions, that describe how the variables in $\mathbf{y}_i(t)$ drift (i.e., change) instantaneously over an infinitely small time interval, as related to their current values, $\mathbf{y}_i(t)$, and $\mathbf{u}_i(t)$, a vector of exogenous predictors (e.g., t). These exogenous predictors are shown here as time-varying, but they may also be time-invariant and specific only to i . $\mathbf{y}_i(t)$ may include derivative variables needed to define higher-order DEs, namely, DEs involving higher-order derivatives than the first derivatives. For instance, a second-order DE model in which the highest-order derivative is the second derivative can be written as two first-order DEs. In this case, $\mathbf{y}_i(t)$ would contain $y_i(t)$ as well as $\frac{dy_i(t)}{dt}$.

If all functions in $\mathbf{f}[\cdot]$ involve only linear functions of $\mathbf{y}_i(t)$, Equation (1) is said to be a linear DE model. Generally, linear functions are those that give a straight line in a graph. A function that is linear in y has the form $f(y) = c + ay$, in which c and a are constants that represent the intercept and slope of the straight line, respectively. A linear DE can — and typically does — have nonlinear integral solutions. These integral solutions are analytic functions that map out all values of $\mathbf{y}_i(t)$ at any $t > 0$ beginning from some initial conditions that specify the values of $\mathbf{y}_i(0)$, namely, the system's values before any changes are realized (Arnold, 1974; Zill, 1993).¹ Linear DEs typically have known analytic solutions.

In contrast, functions that do not fall into any special cases of the form $c + ay$ are generally nonlinear functions. Examples of nonlinear functions include interactions among the variables in $\mathbf{y}_i(t)$, or polynomial functions of $\mathbf{y}_i(t)$ beyond the first degree (e.g., quadratic, cubic functions; Zill, 1993). Some examples of nonlinear DEs have already been proposed in studies of ovulatory regulation (Boker, Neale, & Klump, 2014), circadian rhythms (E. N. 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, Ferrer, & Nesselroade, 2007); and sudden transitions in attitudes (van der Maas, Kolstein, & van der Pligt, 2003). Most nonlinear DEs do not have analytic solutions. In such cases, the trajectories of the variables in $\mathbf{y}_i(t)$ can alternatively be mapped out using numerical solvers, which are approaches for computing the predicted numerical values of the system successively in time using the hypothesized DEs at some pre-determined time steps (Press, Teukolsky, Vetterling, & Flannery, 2002).²

Latent Stochastic Differential Equation (SDE) Models

Suppose that the true processes of interest, represented as $\boldsymbol{\eta}_i(t)$, a $w \times 1$ vector of latent variables at time t , are unobserved but may be identified using the variables in $\mathbf{y}_{it}$, which are possibly contaminated by measurement errors. Let's suppose further that our representation of the change mechanisms of the underlying latent processes may also be imperfect, or are influenced by other sources of uncertainty. This scenario requires the use of a latent stochastic DE (SDE) model, expressed as (Arnold, 1974):

$$d\boldsymbol{\eta}_i(t) = \mathbf{f}(\boldsymbol{\eta}_i(t), \mathbf{u}_i(t)) dt + d\mathbf{w}_i(t), \quad (2)$$

¹These solutions may equivalently be expressed in exact discrete time form (Harvey, 2001), which specifies the values of $\mathbf{y}_i(t_{i,j})$ at discrete time point $t_{i,j}$ using the projected values of $\mathbf{y}_i(t_{i,j-1})$ at a previous time point.

²These time steps are distinct from the time intervals of the observed data — the former can be specified to be smaller than the latter to reduce the numerical errors that arise from such approximations.

$\mathbf{f}(\cdot)$ is again, a vector of differentiable linear or possibly nonlinear drift functions that describe the changes in a vector of latent variables, $\boldsymbol{\eta}_i(t)$, over an infinitely small time interval. These drift functions now depend on $\boldsymbol{\eta}_i(t)$ and $\mathbf{u}_i(t)$, a vector of person- and time-specific covariates. As in Equation (1), $\boldsymbol{\eta}_i(t)$ may include latent derivative variables needed to define higher-order SDEs, namely, SDEs involving higher-order derivatives than first derivatives. $\mathbf{w}(t)$ is a vector of Wiener processes, which can be understood as process noises, or dynamic errors, that contribute to the uncertainty of the change mechanisms of $\boldsymbol{\eta}_i(t)$. Wiener processes have been used to characterize, for instance, Brownian motion, the diffusion of minute particles suspended in fluid (R. Brown, 1828). $d\mathbf{w}_i(t)$ denotes differences in the Wiener processes over dt , with a covariance matrix, $\mathbf{Q}$ (also known as the diffusion matrix), whose values depend on dt . When $\mathbf{Q}$ is a null matrix (i.e., there are no process noises in the system), the equation in (2) reduces to a deterministic ordinary differential equation (ODE) model. This is in contrast to situations where $\mathbf{Q}$ is not a null matrix, in which case future values of $\boldsymbol{\eta}_i(t)$ can only be predicted as subjected to the dynamic errors in $\mathbf{w}_i(t)$. That is, the DEs include a stochastic component so that Equation (2) is an SDE model. Note that the instantaneous rate of change of the Wiener process does not exist or is not defined over infinitely small intervals (Arnold, 1974; Molenaar & Newell, 2003). Thus, I express the dynamic model in the differential form in Equation (2), as opposed to the alternative form in which $\frac{d\boldsymbol{\eta}_i(t)}{dt}$ appears on the left-hand side of the equation.

The latent variables in $\boldsymbol{\eta}_i(t_{i,j})$ at discrete time point $t_{i,j}$ are indicated by a $p \times 1$ vector of manifest observations, $\mathbf{y}_i(t_{i,j})$. These observations are assumed to be measured at individual-specific and possibly irregularly spaced time points $t = t_{i,j}$, $j = 1, \dots, T_i$, as are the vector of person- and time-varying covariates, $\mathbf{u}_i(t)$. The vector of manifest observations is linked to the latent variables as

$$\mathbf{y}_i(t_{i,j}) = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i(t_{i,j}) + \mathbf{A}\mathbf{u}_i(t_{i,j}) + \boldsymbol{\epsilon}_i(t_{i,j}), \quad \boldsymbol{\epsilon}_i(t_{i,j}) \sim N(\mathbf{0}, \mathbf{R}). \quad (3)$$

In Equation (3), $t_{i,j}$ denotes the j th observed time point for person i , $\boldsymbol{\tau}$ is $p \times 1$ vector of intercepts, $\boldsymbol{\Lambda}$ is a $p \times w$ factor loading matrix that links the observed variables to the latent variables, and $\mathbf{A}$ is a matrix of regression weights for the covariates in $\mathbf{u}_i(t_{i,j})$ observed at time $t_{i,j}$. Adopting the modeling tradition in the state-space literature (see e.g., Durbin & Koopman, 2001), we assume that all sources of time dependencies of interest are specified as part of the dynamic model in (2) and consequently, $\boldsymbol{\epsilon}_i(t_{i,j})$ is a $p \times 1$ vector of measurement errors assumed to be serially uncorrelated over time and normally distributed with a mean vector of zeros and covariance matrix, $\mathbf{R}$. Equations (2) and (3) represent the *dynamic model* and *measurement model*, respectively, that collectively define a dynamic system. All of the illustrative examples considered in this article are special cases of Equations (2) and (3).

Model Building and Model Exploration

New methodological extensions continue to augment the repertoire of tools for fitting increasingly complex DE models (Beskos, Papaspiliopoulos, & Roberts, 2009; Mbalawata, Särkkä, & Haario, 2013; Ramsay, Hooker, Campbell, & Cao, 2007). Applications and methodological developments that involve fitting linear ordinary and stochastic DE (ODE and SDE, respectively) models have evidenced considerable growth in recent years (Boker & Graham, 1998; Boker et al., 2014; Deboeck, 2010; Oravecz, Tuerlinckx, & Vandekerckhove, 2016; Oud & Jansen, 2000; Trail et al., 2013; Voelkle & Oud, 2013). In contrast, studies on

methods for fitting nonlinear ODEs and SDEs are still nascent, (e.g., Chow et al., 2007; Chow, Lu, Sherwood, & Zhu, 2016; Cobb & Zacks, 1985; Lu, Chow, Sherwood, & Zhu, 2015; Molenaar & Newell, 2003; Singer, 2002; Wagenmakers, Molenaar, Grasman, Hartelman, & van der Maas, 2005), even though a subset of nonlinear DE models have received widespread attention.

Despite recent advances in fitting confirmatory DE models, few theories exist in the social and behavioral sciences to guide the formulation of confirmatory models of change mechanisms. Exploratory approaches can be helpful as a first step to unveil possible determinants of and linkages among change mechanisms when there is a lack of theory to guide confirmatory modeling efforts. Two-stage approaches that produce intermediate output such as derivative estimates can be especially helpful for model exploration and building purposes because the derivative estimates serve as proxies for a system's instantaneous changes and any higher-order changes thereof (Butner, Gagnon, Geuss, Lessard, & Story, 2015; Deboeck, Montpetit, Bergeman, & Boker, 2009). In these approaches, researchers first obtain derivative estimates by using different variations of numerical differencing procedures to approximate the instantaneous (first and higher-order) changes, and subsequently use the derivative estimates for model building and exploration purposes. Model building then becomes a direct *variable selection* problem wherein the goal is to identify predictors that can help explain those derivative estimates. The utility of these derivative estimation approaches can be further expanded by capitalizing on the recent surge of nonparametric (e.g., spline-based) modeling tools that allow researchers to explore, without assuming any a priori model, interrelations among variables and their associated derivatives. These spline-based methods include GAMs (GAMs; Hastie & Tibshirani, 1990), modifications of classical GAMs to enable efficient variable selection from a large pool of candidate predictors (Hastie, Tibshirani, & Friedman, 2009; S. N. Wood, 2006), tools that target broad accessibility (Li, Tan, Huang, Wagner, & Yang, 2014), as well as functional data analysis (FDA) tools that offer flexible, spline-based approximations for curves, their derivatives, and corresponding analysis (Ramsay & Silverman, 2005).

For the purpose of derivative estimation, the generalized local linear approximation (GLLA; Boker, Deboeck, Edler, & Keel, 2010) and the generalized orthogonal local derivative (GOLD; Deboeck, 2010) are among some of the better-known tools in the psychological literature. Chow, Bendezú, Cole, and Ram (2016) provided an overview of the respective strengths of the GLLA, GOLD, and a spline-based approach typically adopted in the FDA literature. Particular advantages of the FDA approach include built-in mechanism to accommodate irregularly spaced data, and enhanced smoothing of the derivative estimates in cases involving noisy, possibly nonlinear dynamics. Here, I utilize FDA for derivative estimation in all illustrative examples. The key ideas behind the FDA approach to derivative estimation are outlined next.

Spline-Based FDA

FDA is a branch of analytic methods focusing on the analysis of curves, within which splines and derivatives play a prominent role. One type of popular spline function is the basis spline (B-spline; De Boor, 1977, 1978; Dierckx, 1993), which approximates a time series at any

arbitrary time t as a linear combination of B basis functions, $\phi_{b,i}(t)$, as:

$$\eta_i(t) \approx \hat{\eta}_i(t) = \sum_{b=0}^B c_{b,i} \phi_{b,i}(t), \quad (4)$$

where $\phi_{b,i}(t)$ is the b th basis function for person i , and $c_{b,i}$ is its associated weight or coefficient. All of the B basis functions are known and fixed functions of t , usually taken to be some form of polynomials functions up to degree B , one popular special case of which is the B -splines. B -splines serve to approximate segments of a time series in a piecewise way using polynomials (De Boor, 1977; Dierckx, 1993). Each segment of the time series is separated from its immediately adjacent segment(s) by a *knot point*, with the first and last measurement occasions typically constituting the outer or ending knots. The number of basis functions used in each segment, also known as the order of a B -spline, is equal to $B + 1$. 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.

To ensure that smoothness of the approximation curve at each interior knot point, two adjacent polynomials are typically specified to match in the values of a fixed number of their derivatives, usually chosen to be $B-1$. Using this convention, a spline of degree 0 yields a step function that is discontinuous at knots; a spline of degree 2 is piecewise quadratic with matching level and first derivative at the interior knot points. Cubic spline – a popular spline function in many substantive applications (e.g., Tarvainen, Georgiadis, Ranta-aho, & Karjalainen, 2006) – is piecewise cubic with matched 1st and 2nd derivatives at knot points to yield visibly smooth approximation curves.

Once $\hat{\eta}_i(t)$ is available from Equation (4), the derivatives of $\hat{\eta}_i(t)$ of order p then follow directly from differentiating Equation (4) with respect to time, which yields:

$$\frac{d^p \hat{\eta}_i(t)}{dt^p} = \sum_{b=0}^B c_b \frac{d^p \phi_{b,i}(t)}{dt^p}. \quad (5)$$

To proceed with derivative estimation, one requisite step is to estimate the unknown coefficients, c_b . However, if too many values of c_b are allowed to be non-zero, one may end up with approximation curves, $\hat{\eta}_i(t)$ and corresponding derivative estimates that are overly “wiggly” — in other words, they capture too much of the nuanced fluctuations in the data. Thus, it is often of interest to use some regularization procedures to penalize against excessive roughness in the approximation curves to ensure that they satisfy some notion of smoothness (Ramsay & Silverman, 2005). One way of doing so is to estimate the basis function coefficients, c_b , by minimizing the penalized residual sum of squares function

$$PENSSSE_\lambda = \sum_{i,j} [\eta_i(t_{i,j}) - \hat{\eta}_i(t_{i,j})]^2 + \lambda PENALTY(\hat{\eta}_i) \quad (6)$$

where $PENALTY(\hat{\eta}_i)$ is a penalty function that captures the extent of deviations from a predefined smoothness criterion; $\lambda \geq 0$ is a smoothing parameter such that the larger λ is, the heavier the penalty (i.e., the estimated curve is smoother). λ has to be estimated, or selected using selection criteria such as the generalized cross-validation index (GCV; Craven & Wahba, 1978) or information criterion measures (Tan, Shiyko, Li, Li, & Dierker, 2012).

In this article, I use the GCV for λ selection purposes, with GCV values being preferred as they indicate better cross-validation results.³

Different penalty functions may be used to regularize the approximation curves and their corresponding derivatives. One penalty function typically used in the context of derivative estimation purposes is

$$PENALTY(\hat{\eta}_i) = \int_{\Gamma} \left(\frac{d^{m+2}\hat{\eta}_i(s)}{ds^{m+2}} \right)^2 ds, \quad (7)$$

where m is the highest derivative desired, and $\int_{\Gamma}$ denotes integration (i.e., “summation”) over Γ , a bounded interval containing the range of t values that are of interest to the researcher. For instance, with $m = 2$ (i.e., a researcher wishes to estimate up to the second derivatives), the integrated squared fourth derivative is used in the penalty function (7) to penalize against excessive curvature in the second derivatives. Consequently, B-splines of order 6 (two higher than the order of the derivatives invoked in the estimation process) may be used for estimation purposes.

The number and placement of knot points are among some of properties that determine characteristics of the approximated curves. Even though some automated schemes exist to help guide these decisions (e.g., Eilers & Marx, 1996; Tan et al., 2012; S. N. Wood, 2003), simultaneous estimation of the smoothing parameter and other quantities related to the knot-point is often computationally formidable. In all the illustrative examples in this article, I place the knot points, a priori, at the observed measurement occasions and choose the smoothing parameter, λ , by minimizing the GCV function. Then, 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 (6) — (7). Estimation of the basis coefficients and smoothing parameter may be repeated, as needed, until reasonable approximations of individual curves are obtained, for instance, through graphical inspection of the approximation curves. Other variations to this approach to accommodate higher degrees of heterogeneity across curves are described in the Other Practical Issues section.

Graphical and Other Variable Selection Tools

Many of the initial hurdles to model building may be greatly circumvented by having accessible graphical tools that can help highlight and clarify the relations among the variables that define a dynamical system, especially for the derivative variables (Butner et al., 2015). In this section, I describe selected standard regression diagnostic and variable selection tools that can be used to detect interrelations among variables and/or their derivatives once derivative estimates of reasonable quality have been computed.

Graphical tools. Many popular exploratory tools within the realm of dynamical systems analysis are graphical tools that help to highlight the topological properties of various dynamical systems. Examples of such plots include phase portraits, which are plots of the variables (including derivative variables) that constitute a dynamical system (see e.g.,

³In a standard leave-out-one cross-validation approach, the goal is to optimize fit by minimizing the sum of the squared discrepancies from predicting each observed data point using an approximation curve constructed using coefficients estimated using all but that specific data point. The GCV generalizes this kind of leave-one-out approaches by incorporating a weight function to accommodate scenarios with irregularly spaced data points and non-periodic curves (Craven & Wahba, 1978).

Butner et al., 2015); and vector field plots, which show the predicted (model-implied) changes characterizing a dynamical system starting from one or multiple initial conditions (for a statistically enhanced version of the vector field plot see Boker & McArdle, 1995). When used with noise-free theoretical dynamical systems, these plots often show clearly discernible patterns of regularity in dynamics that have become the “signatures” of various dynamical systems. When used with empirical data that are contaminated with noise, the regularity in dynamics as revealed through the plots may not be immediately salient. Thus, additional tools are needed to help clarify the interrelations among variables.

Many graphical tools from the regression framework have considerable potential to facilitate model building, but are clearly underutilized in the dynamical systems literature. One example includes variable selection tools that are often utilized in the regression framework to help researchers select the predictors (variables) that can best explain one or more dependent variables in some optimal but parsimonious way. In a similar vein, these tools can also be used to help build dynamical system models, except that some of the implicated variables are now derivative variables. One such graphical variable selection tools in the regression framework is the component-plus-residual plot (Fox, 2015; F. S. Wood, 1973). A component-plus-residual plot is a plot of the residuals of the dependent variable against a predictor after the effects of other predictors have been partialled out. Additionally, a loess line and a linear least squares line are overlaid on the plot to help visualize any possible deviations in associations from linearity (Fox, 2015). This plot can be created, for instance, using the *crPlots* function in the “*car*” library in R (Fox & Weisberg, 2019). In this article, I will demonstrate the use of components-plus-residual plots with other commonly adopted plots for visualizing higher-dimensional data, such as contour plots, to visualize (possibly nonlinear) patterns of association among derivative and level variables.

Semiparametric GAM-based tools. In the absence of clear theories of change, one alternative route is to begin from a reasonably flexible model — specifically, a semiparametric model — that explores, within a partially confirmatory framework, the unknown interrelations among a set of level and derivative variables, as well as their influences on one or more dependent variables of interest. For instance, one possible semiparametric model shown in the special case involving two arbitrary predictors, $u_{1,i}$ and $u_{2,i}$, on an *observed* dependent variable y_i , can be expressed as (Harrell, 2001; Hastie & Tibshirani, 1990; S. N. Wood, 2003):

$$\begin{aligned}
 E(y_i|u_{1,i}, u_{2,i}, \theta_i) &= g(u_{1,i}, u_{2,i}, \theta_i) && \text{parametric component} \\
 &+ s_1(u_{1,i}) + s_2(u_{2,i}) && \text{additive smoothed effects} \\
 &+ s_3(u_{1,i})u_{2,i} + s_4(u_{2,i})u_{1,i} && \text{varying coefficients} \\
 &+ s_5(u_{1,i}, u_{2,i}) && \text{tensor products for jointly nonlinear effects,} \quad (8)
 \end{aligned}$$

wherein $g(\cdot)$ denotes a parametric function (of known — usually linear — form) involving the two predictors and θ_i , a vector of person-specific parameters. Following this component are five nonparametric functions (s_1 — s_5) that have unknown forms, which are each approximated via regularized spline functions. The notation $s_k(u_{h,i})$ denotes the k th nonparametric function that involves the h th predictor, $u_{h,i}$; whereas $s_k(u_{h,i})u_{l,i}$ denotes multiplication of the predictor $u_{l,i}$ with the k th nonparametric function that involves the predictor $u_{h,i}$.

The nonparametric functions in Equation (8) include three sets of terms, which I will elaborate in turn. The first set includes functions that capture additive effects of each

predictor ($s_1 - s_2$). As an example, consider the Yerkes-Dodson law (Yerkes & Dodson, 1908), which predicts an inverted U-shaped relation between individuals' arousal levels and their performance on difficult cognitive tasks. That is, moderate, as opposed to low or high arousal levels, are assumed to optimize performance. Suppose a researcher wishes to clarify this relation empirically, as opposed to imposing a parametric (i.e., known mathematical) function that dictates the nature of this relation. In this case, the researcher may specify performance to be the dependent variable, y_i , and arousal level to be the predictor, $u_{1,i}$. Consequently, any linear or nonlinear effect of arousal level on performance may be revealed through the estimated form of s_1 .

The second set of terms ($s_3 - s_4$) are varying coefficient functions (Hastie & Tibshirani, 1993) that allow the effect of each predictor to vary in a linear or nonlinear manner depending on the values of another predictor. Building on the same example, the Yerkes-Dodson law additionally states that the inverted U-shaped relation is expected to arise under challenging cognitive tasks. With simpler tasks, individuals' performance levels are expected to increase consistently with rises in arousal levels. In this case, the researcher may specify task difficulty level to be $u_{1,i}$, arousal level to be $u_{2,i}$, and use the term $s_3(u_{1,i})u_{2,i}$ to approximate the effect of arousal level as dependent on task difficulty level. Even though I use the example of a categorical predictor for $u_{1,i}$ here, $u_{1,i}$ may also be continuous in nature. In the case where $u_{1,i}$ represents time, the resultant function $s_3(u_{1,i})u_{2,i}$ shows the time-varying effect of $u_{2,i}$ on y_i . Such time-varying coefficient models have been applied in areas such as psychophysiology (Tarvainen et al., 2006), brain imaging (Molenaar, Beltz, Gates, & Wilson, 2007), and affect (Chow, Zu, Shifren, & Zhang, 2011).

The third set of terms in Equation (8) consists of a tensor product function, $s_5(u_{1,i}, u_{2,i})$, that allows for approximations of jointly nonlinear effects involving both $u_{1,i}$ and $u_{2,i}$. Roughly speaking, tensors are multidimensional arrays that comprise a series of univariate functions. In our case, these univariate functions are univariate spline functions. Tensor products are used to approximate possibly nonlinear multivariate functions via linear combinations of the products of all the univariate splines. Extending the earlier example, suppose the relations among arousal level, task performance, and task difficulty level are more complex than was originally hypothesized by the Yerkes-Dodson law. Specifically, the researcher speculates that task difficulty itself affects task performance following a “Z-shaped” function: that is, relatively little decrements in task performance are expected as the task varies from easy to moderate difficulty, followed by precipitous declines in performance when the task moves beyond some threshold difficulty level. In this case, task difficulty itself has a nonlinear relation to task performance. Thus, the interactive relation between task difficulty and arousal level is best clarified with a tensor product term.

The full model, which involves a combination of parametric and nonparametric components, is thus a semiparametric additive model. If the data for model fitting follow any of the special cases in the exponential family (e.g., Gaussian, Poisson, Bernolli, Binomial, Multinomial), then $E(y_i|\cdot)$ can be mapped to the observed data y via a link function, thereby constituting a generalized linear modeling framework (McCullagh & Nelder, 1989).

Extrapolating to the context of DE modeling, the dependent variable is typically the highest-order derivative of interest to a researcher, the m th order derivative at time $t_{i,j}$. Furthermore, $u_{1,i}$ and $u_{2,i}$ may be level or derivative variables of a lower order than m . Here, an arbitrary example involving two predictors is shown. Scenarios that motivate the

use of such semiparametric variable selection routines in the regression framework typically involve a large number of potential predictors. The form of the relation of each predictor to the dependent variable is unknown and may be linear or nonlinear in nature. The use of penalized estimation routines further affords the possibility to shrink the coefficients associated with unimportant predictors to zero, thereby accomplishing variable selection and model comparison simultaneously (Geweke, 1996; Lu, Chow, & Loken, 2017).

I use the penalized least squares estimation routines in the R package, Mixed GAM Computation Vehicle with Automatic Smoothness Estimation (mgcv; S. Wood, 2018), to perform estimation of GAMs. The thin plate regression splines use an eigenvalue decomposition procedure to select piecewise regression spline coefficients that can maximize the amount of variance explained in the data (S. N. Wood, 2003, 2006). For didactic introductions to GAMs see McKeown and Sneddon (2014) and S. N. Wood (2006).

Single-Stage, Confirmatory Approaches to Fitting SDE Models: The Continuous-Discrete Extended Kalman Filter (CDEKF) as One Possible Approach

I have highlighted some possible tools for exploring and building DE models, both graphically and via semiparametric modeling techniques. However, as I will demonstrate in one of the illustrative examples, multi-stage exploratory/semiparametric approaches, though flexible, can often come at the costs of reduced statistical precision, efficiency, and power compared to single-stage, confirmatory approaches that fit correctly specified models to the data. Thus, for inferential purposes, it can be advantageous to supplement initial exploratory results with results from single-stage, confirmatory model fitting.

Some limited confirmatory tools have been proposed and developed in the statistical and psychometric literature for fitting latent SDE models (e.g., Chow et al., 2007; Chow, Lu, et al., 2016; Driver, Oud, & Voelkle, 2017; Kou, Olding, Lysy, & Liu, 2012; Lu et al., 2015; Molenaar & Newell, 2003; Oravecz et al., 2016; Singer, 2010, 2012). Of these approaches, I use the *continuous-discrete time extended Kalman filter* (CDEKF) algorithm, provided as part of the R package, *Dynamic Modeling in R* (*dynr*; Ou, Hunter, & Chow, 2018, revised and resubmitted), for confirmatory estimation of linear and nonlinear DE models with Gaussian data.

The implementation of the CDEKF has been described in more detail elsewhere (Chow et al., 2018; Kulikov & Kulikova, 2014; Kulikova & Kulikov, 2014) and is not reiterated here. In brief, the CDEKF (Bar-Shalom, Li, & Kirubarajan, 2001; Kulikov & Kulikova, 2014; Kulikova & Kulikov, 2014) is a procedure for estimating the latent variables that appear in a system of (linear and possibly nonlinear) SDEs. The linear, discrete-time analogue of the CDEKF, known as the linear Kalman filter, has some known parallels to well-known factor score estimators in the psychometric literature (Chow, Ho, Hamaker, & Dolan, 2010; Dolan & Molenaar, 1991; Lawley & Maxwell, 1973; Oud, van den Bercken, & Essers, 1990). The CDEKF algorithm as implemented in the *dynr* package uses the fourth-order Runge-Kutta (Press et al., 2002), one possible numerical DE solver that derives successive approximations of the over-time values of the latent variables via weighted averages of four sets of model-implied changes. In cases involving nonlinear DEs, a Jacobian matrix composed of first-order symbolic differentiations of the (possibly) nonlinear dynamic functions with respect to the latent variables is used in all covariance functions to enable approximations of the nonlinear changes using first-order Taylor series expansion (i.e., piecewise linear ap-

proximations).

Under the constraint of a linear measurement model and normally distributed process noises, Chow and colleagues (2007) noted that a closed-form log-likelihood function can be constructed using by-products of Kalman filtering algorithms, and maximized using an optimization procedure (e.g., Newton Raphson) to obtain estimates of all unknown parameters in the system. The estimates of the standard error are obtained using a matrix of numerical Hessians (second derivatives) of the loglikelihood function with respect to the parameters. In addition, information criterion measures such as the Akaike Information Criterion (Akaike, 1973) and Bayesian Information Criterion (Schwarz, 1978) can also be computed based on the loglikelihood function (Harvey, 2001) for model comparison purposes. Thus, the procedures for confirmatory model fitting adopted in this article capitalize on a suite of routines implemented in *dynr* to handle latent variable estimation, parameter estimation, and model comparisons involving ODEs/SDEs.

Illustrative Examples

I present four illustrative examples to demonstrate the respective strengths of multistage exploratory vs. single-state confirmatory approaches. I begin with a univariate ODE model (Illustration I) and a bivariate ODE model (Illustration II) that are relatively well-known in the psychometric literature. These illustrations are followed by two illustrations (III and IV) that utilize exploratory tools to reveal evidence of qualitative changes in dynamical systems. The last illustration (V) underscores some of the inadequacies of two-stage exploratory approaches in comparison to results from single-stage confirmatory modeling. For all illustrations, I used numerical solvers to generate all the data; the time intervals for deducing numerical solutions were specified to coincide with the empirically observed time intervals. The code for all illustrative examples is provided in the supplementary materials.

Illustration I: Linear Oscillator Model

My first illustrative example features the simpler case of a linear ODE with only measurement errors and no process noise. In this illustration, standard derivative estimation approaches such as the FDA, the GLLA and GOLD typically provide satisfactory derivative estimates; subsequent use of these derivative estimates for model fitting in a second stage generally yields reasonable estimates of the parameters and their standard errors. This example serves to demonstrate the utility of component-plus-residual plots in revealing interrelations among level and derivative variables from FDA.

One of the most prominent and broadly used ODE in the psychological literature is the damped linear oscillator model (Boker & Graham, 1998), a mathematical model describing the behavior of a swinging pendulum with friction. This model is expressed as:

$$\frac{d^2\eta_i(t)}{dt^2} = \omega\eta_i(t) + \zeta\frac{d\eta_i(t)}{dt}, \quad (9)$$

in which $\eta_i(t)$ represents the true level or displacement of the pendulum relative to its equilibrium position (the center of motion); ω 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. When $\zeta < 0$, shows damping, or reduction in the magnitudes of displacement over time. Alternatively, if $\zeta > 0$, the extent to which the pendulum shows amplification in displacement over time. Damping and amplification are both characteristics that pertain to how a system's amplitude of

fluctuations changes over time. With damping and in the absence of further external shocks, the system will, as time increases, eventually settle into a stable fixed point — often called attractor of sink — at 0. With amplification, 0 becomes an unstable fixed point, termed a repeller, from which the system moves away as time increases.

Equation (9) is the DE representation of a swinging pendulum that, when subjected to friction, shows damping or successive reductions in its displacements from a set-point, eventually coming to rest at the set-point. It also describes how a heater, given input from a thermostat, operates to reduce the discrepancies between the current room temperature and a target temperature. Chow, Ram, Boker, Fujita, and Clore (2005) used this mathematical model to describe an “emotional thermostat” that delineates emotion regulation as a process through which individuals damp deviations in their emotional levels toward their characteristic set point levels.

I generated over-time trajectories of $\eta_i(t)$ using Equation (9) with $\omega = -0.8$, $\zeta = -0.1$ for $n = 50$ individuals from $t = 0$ to 10 across $T_i = T = 100$ measured time points. The initial conditions, $\eta_i(0)$ and $\frac{d\eta_i(0)}{dt}$, were specified to be normally distributed across people with zero means and variances of 4 and 0.25, respectively. The time intervals between successive time points, namely, $\Delta(t_{i,j}) = t_{i,j} - t_{i,j-1}$, were fixed to 0.1 for all of the $j = 1, \dots, T_i$ measurement occasions, and for all $i = 1, \dots, n$ individuals. In addition, after obtaining the numerical solutions, normally distributed measurement errors with zero mean and a variance of 1.0 were added to $\eta_i(t)$ to yield $y_i(t)$.

As noted, the over-time trajectories of $\eta_i(t)$ (see Figure 1) were generated with the same set of values for ω and ζ . Differences in the magnitudes of fluctuations over time were due primarily to individual differences in the initial conditions. The lack of individual differences in dynamics (i.e., in ω and ζ) may be difficult to deduce from the over-time individual trajectories in Figure 1(A) per se, but inspection of the corresponding component-plus-residual plots in Figure 1(B)—(C) helps clarify the lack of salient interindividual differences in the slope relating $\eta_i(t)$ to $\frac{d^2\eta_i(t)}{dt^2}$ after the effect of $\frac{d\eta_i(t)}{dt}$ has been partialled out. In addition, the component-plus-residual plots indicate that it is reasonable to assume linear relations among the levels and first derivatives with the second derivatives, as evidenced by the clear overlap between the linear regression lines (the dashed lines) and the loess lines (the solid lines). Unlike the linear regression line, the loess lines do not impose linearity assumptions on the relations between the independent and the dependent variables, and thus, has been used as part of the diagnostic steps for identifying potential nonlinear relations in a regression context.

In this illustrative example, given the linear relations among all independent and dependent variables and the lack of interindividual differences in dynamics, the smoothed levels as well as first and second derivative estimates can be used as variables in a standard regression model. The estimated regression coefficient linking $\hat{\eta}_i(t)$ to $\frac{d^2\hat{\eta}_i(t)}{dt^2}$ provides an estimate of ω , and the coefficient linking $\frac{d\hat{\eta}_i(t)}{dt}$ to $\frac{d^2\hat{\eta}_i(t)}{dt^2}$ provides an estimate of ζ . Provided that the data have reasonable reliability and sample size, and the smoothed levels offer a reasonable approximation of the underlying true scores, ω and ζ can generally be well recovered. In previous work, my collaborators and I have tried scenarios where reliability, defined as the ratio between the true score and total variance, was as low as .6 and the true parameters were still well recovered. In the illustration shown here, the reliability of

the simulated data was around .57. Of course, myriad other factors may also impact the quality of the derivative estimates, such as the complexity of the DE models, the presence of process noises, and the presence of other confounds, such as under- or over-smoothing in the estimated levels and derivatives used for model fitting. Some of these factors are explored in subsequent illustrative examples.

Illustration II: Classical Predator-Prey Model

In this example, I utilize a nonlinear ODE model, the classical predator-prey model (Lotka, 1925; Volterra, 1926), often termed the Lotka–Volterra equation, to demonstrate how non-linear relations among level and derivative variables may be visualized and explored using graphical methods, GAMs, and another common technique to probe interaction effects in regression analysis, simple slope analysis (Cohen, Cohen, West, & Aiken, n.d.).

The classical predator-prey model captures the interaction between a predator and a prey population as:

$$\frac{d\eta_1(t)}{dt} = r_1\eta_1(t) - a_{12}\eta_1(t)\eta_2(t), \text{ and} \quad (10)$$

$$\frac{d\eta_2(t)}{dt} = -r_2\eta_2(t) + a_{21}\eta_1(t)\eta_2(t), \quad (11)$$

where $\eta_1(t)$ corresponds to the true density of the prey population at time t and $\eta_2(t)$ the true density of the predator population. The terms $\frac{d\eta_1(t)}{dt}$ and $\frac{d\eta_2(t)}{dt}$ on the left-hand-side of the two equations represent the rates of change in the densities of the prey and predator populations at time t . The parameter $r_1 > 0$ is used to represent the growth rate of the prey population in the absence of the predator population (i.e., when $\eta_2(t) = 0$); and $r_2 > 0$ is the death rate of the predator population in the absence of its sole food source (i.e., the prey, when $\eta_1(t) = 0$). Relatedly, the interaction between the predator and prey population is hypothesized to lead to negative outcome for the prey population (thus Equation (11) has the component $-a_{12}$ instead of $+a_{12}$), whereas the predator population is assumed to benefit from this interaction (the magnitude of which is determined by a_{21}). In certain parameter ranges, this model is known to yield cyclic fluctuations in predator and prey densities in a lead-lag manner.

The classical predator-prey model features a single population of predator and prey; thus, there is no subject or population index i in Equation (11). Chow et al. (2007) used a multiple-subject extension of the predator-prey model to describe the “encroachment” of positive affect of individuals in a dyadic romantic relationship by the negative affect of their partners. An extended version of this model was used by Chow and Nesselroade (2004) to represent age differences in cognitive performance due to individuals’ increased difficulty in ignoring interference from irrelevant information with age. In other words, the irrelevant information is “preying on” individuals’ ability to attend to cognitive tasks. Here, I generated simulated trajectories for $n = 10$ participants across $T = 200$ time points with 0.08 as the time interval. The parameters were set to be $r_1 = 1.5$, $r_2 = 1$, $a_{12} = .5$, $a_{21} = .4$, with normally distributed additive measurement errors with means of zero and variances of 1.0 for both the prey and the predator. Note that because of the specification of normally distributed measurement errors, the density values of prey and predator could actually extend below 0 – an artifact that does not make sense from a population density standpoint. To aid interpretation, these densities were subjected to exponential transformations so that

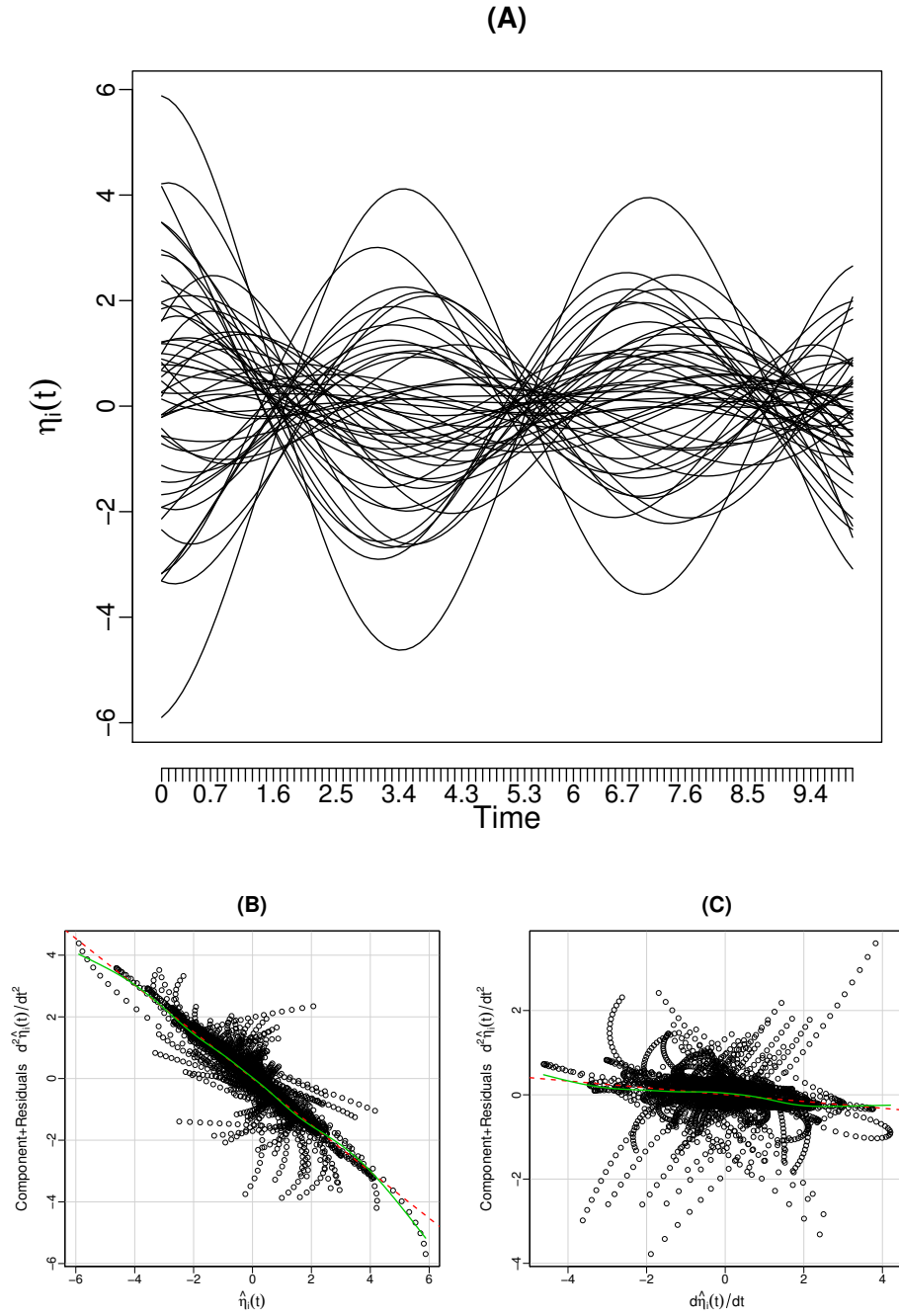


Figure 1. (A) Over-time trajectories generated using the linear oscillator model; (B) component-plus-residual plot revealing the association between $d^2\hat{\eta}_i(t)/dt^2$ and $\hat{\eta}_i(t)$ after the linear effect of $d\hat{\eta}_i(t)/dt$ has been partialled out; (C) component-plus-residual plot revealing the association between $d^2\hat{\eta}_i(t)/dt^2$ and $d\hat{\eta}_i(t)/dt$ after the linear effect of $\hat{\eta}_i(t)$ has been partialled out.

the final observed data, $y_{1i}(t)$ and $y_{2i}(t)$, could only take on values in the positive range. These transformations are similar to those used in Poisson regression models, which posit that the logarithm of the expected value of a dependent variable (usually some sort of count data) is a linear combination function of a set of predictors. In other words, the expected value of the dependent variable is explained by exponential transformation of the linear combination of predictors (Fox, 2015; McCullagh & Nelder, 1989).

Simulated over-time trajectories generated using the specification above are plotted in Figure 2(A). Here, the classical cyclic lead-lag relations between the predator and prey populations, following the exponential transformations, are manifested as sequential bursts in density values. The corresponding scatterplot of the observed and smoothed log transformed prey and predator density values, as shown in Figure 2(B), provides an alternative portrayal of the cyclic dependencies between the two species. The scatterplot indicates that there is a positive relation between the smoothed log density values of the two species, but only up until particular levels of smoothed log predator density (around 4 – 6). Above this value, the predator’s smoothed log density either stays stagnant, or actually declines with further increases in smoothed log prey density. This relationship corresponds to the delay with which the predator’s density resumes growth following the earlier depletion in the prey’s density. The scatterplot depicts the relation between two of the key variables in the predator-prey system (levels of the two processes), and as such, is one example of a phase portrait. Note, however, that when measurement errors are present, the cyclic relations between the two sets of observed log density values are not evident without smoothing from FDA.

The component-plus-residual plots shown in Figures 2(C) and (D) feature residuals from two linear regression models in which the smoothed first derivatives of the log prey and predator density values, denoted respectively as $d\log(\hat{\eta}_{1i}(t))/dt$, and $d\log(\hat{\eta}_{2i}(t))/dt$, were predicted using only the linear effects of the two species’ smoothed log density values, denoted respectively as $\log(\hat{\eta}_{1i}(t))$ and $\log(\hat{\eta}_{2i}(t))$. In this case, reliance on the loess lines in the component-plus-residual plots alone did not provide enough sensitivity to detect the nonlinear dependencies between the two species: There were very little deviation of the loess lines from the linear regression lines in Figures 2(C) and (D). However, some indication of potential nonlinearity can be deduced from the divergence in the values of the smoothed first derivative estimates at the same values of $\hat{\eta}_{1i}(t)$. Such divergence is more salient in Figure 2(C), in which values of the first derivatives can be both positive and negative at one particular set values of smoothed, log transformed prey density values.

To probe for potential nonlinear dependencies between the two sets of derivative variables, I fit a GAM in which the linear parametric effects of $\hat{\eta}_{1i}(t)$ and $\hat{\eta}_{2i}(t)$, as well as the tensor product between the two (i.e., the term, $s_5(\hat{\eta}_{1i}(t), \hat{\eta}_{2i}(t))$ in Equation (8), with $u_{1,i}$ and $u_{2,i}$ set to $\hat{\eta}_{1i}(t)$ and $\hat{\eta}_{2i}(t)$, respectively), were used to predict the smoothed first derivative estimates of the log predator density, $d\log(\hat{\eta}_{2i}(t))/dt$. Results from model fitting indicated that the linear parametric effects as well as the tensor product term were all statistically significant, suggesting that the linear terms alone did not adequately characterize the patterns of the smoothed derivative estimates.

One way to visualize the effects of the model is to use a contour plot (see Figure 3(A)). In the contour plot in Figure 3(A), the vertical axis represents smoothed log predator density level, $\log(\hat{\eta}_{2i}(t))$, and the horizontal axis plots values of the smoothed log prey density level,

$\log(\hat{\eta}_{1i}(t))$. Lines on the plot are known as *contour lines*. They are marked with numbers that convey the predicted values of the dependent variable in the GAM of choice, in this case the smoothed first derivative of the log transformed predator density, $\frac{d\log(\hat{\eta}_{2i}(t))}{dt}$. Each contour line connects values of the predictors on the horizontal and vertical axes that yield the same predicted value for the dependent variable. In instances where the effects of the two plotted independent variables on the dependent variable are linear and do not depend on the values of the other variable, the contour lines would be roughly evenly spaced across all values of those independent variables.

Here, we observe that the contour lines at very low log predator density values (e.g., $\log(\hat{\eta}_{2i}(t)) < 1.5$) were closer together than those at high smoothed log predator density values. Within this region of the contour plot, even slight changes in log prey density are predicted to yield relatively large changes in the values of growth (i.e., positive predicted first derivatives) or declines (negative first derivatives) in log predator density. In contrast, the contour lines were further apart at high values of log predator density. This suggests that at high values of log predator density, greater increases in log prey density are needed to stop/reduce the declines in log predator density when the levels of log prey density are low; or further increase the growth in log predator density when the levels of log prey density are high. In sum, the growth in log predator density with increase in log prey density does not occur at a uniform rate but rather, depends on the current log density size of the predator. Thus, even without articulating a full parametric model, some of the interrelations between the two species may be inferred from the contour plot.

Finally, I show how standard procedures in classical regression analysis, such as simple slope analysis used to probe interaction effects between predictors (Aiken & West, 1991), may be utilized to facilitate interpretations of the nature of the predator-prey interaction here. Here, I probed the simple effect of the smoothed log predator density, $\log(\hat{\eta}_{2i}(t))$ on the log prey population's first derivatives, $d\log(\hat{\eta}_{1i}(t))/dt$, at different values of smoothed prey density values, $\log(\hat{\eta}_{1i}(t))$, using a linear regression model in which $d\log(\hat{\eta}_{1i}(t))/dt$ was predicted using linear parametric effects of $\log(\hat{\eta}_{1i}(t))$, $\log(\hat{\eta}_{2i}(t))$, and their product term, $\log(\hat{\eta}_{1i}(t))\log(\hat{\eta}_{2i}(t))$. In Figure 3(B) in which the simple slope of the smoothed log predator density, $\log(\hat{\eta}_{2i}(t))$, is depicted on the vertical axis (marked as θ), the simple slope values are observed to be statistically significant and positive at smoothed log prey density values that are greater than 2.57, but statistically significant and negative at log prey density values that are less than 2.40. The switch in values of the simple slopes from negative to positive provides some indication of how the same unit of increase in log predator density was associated with reductions as compared to increases in log predator density levels at low versus high log prey density level.

Illustration III: Bifurcation as Discontinuous Changes in Dynamics

One of the ways in which a dynamical system can show discontinuous changes in dynamics is through the phenomenon of bifurcation. Bifurcation refers to qualitative changes in the dynamics of a system with continuous changes in one or more parameters in the system (Poston & Stewart, 1978). One everyday example of bifurcation resides in instances where individuals show sudden switch from walking to running on a treadmill as the treadmill slowly speeds up or slows down. In this case, treadmill speed is a control variable — or specifically, a bifurcation variable, rather than a parameter. Continuous changes in the

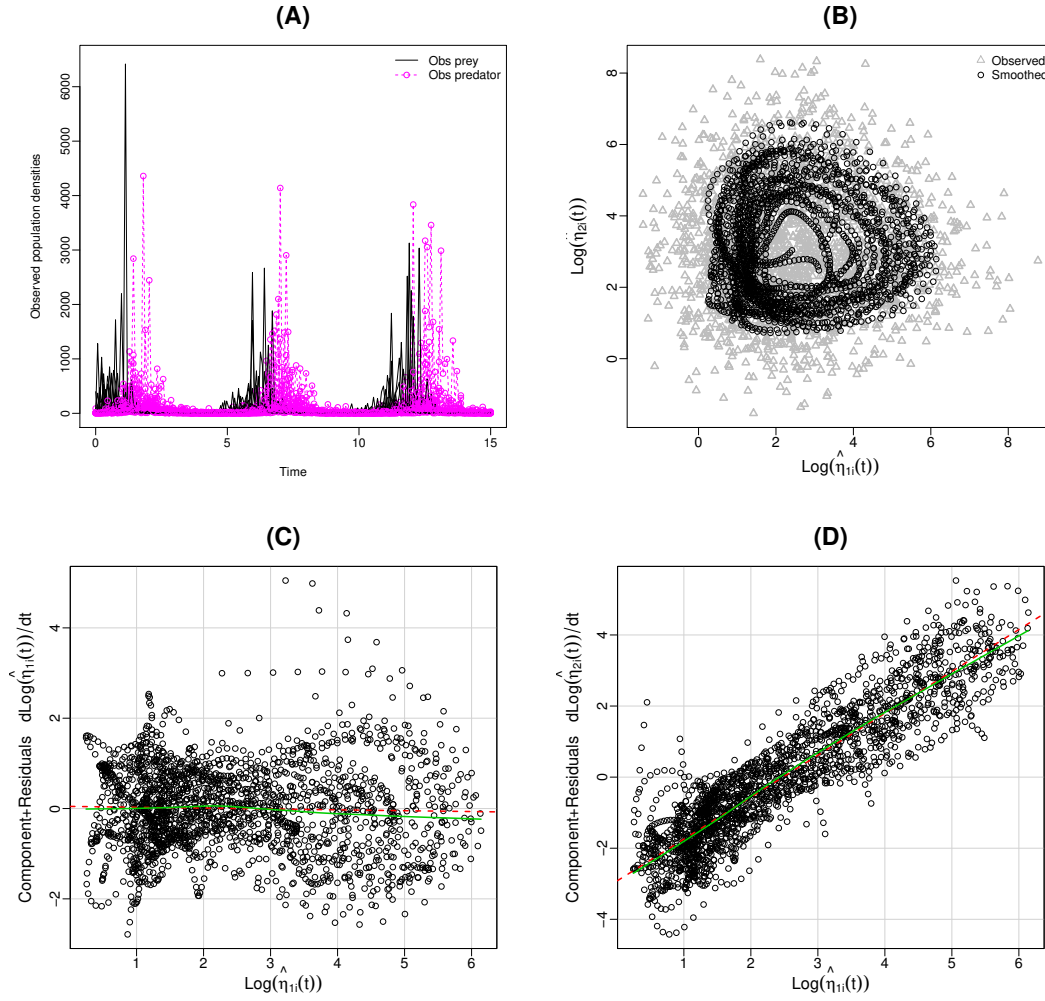


Figure 2. (A) Over-time trajectories generated using the classical predator-prey model; (B) a scatterplot of the smoothed predator density levels, $\hat{\eta}_{2i}(t)$, against the smoothed prey density levels, $\hat{\eta}_{1i}(t)$, overlaid with their corresponding observed values; (C) component-plus-residual plot revealing the association between $d\hat{\eta}_{1i}(t)/dt$ and $\hat{\eta}_{1i}(t)$ after the linear effect of $\hat{\eta}_{2i}(t)$ has been partialled out; (D) component-plus-residual plot revealing the association between $d\hat{\eta}_{2i}(t)/dt$ and $\hat{\eta}_{1i}(t)$ after the linear effect of $\hat{\eta}_{2i}(t)$ has been partialled out

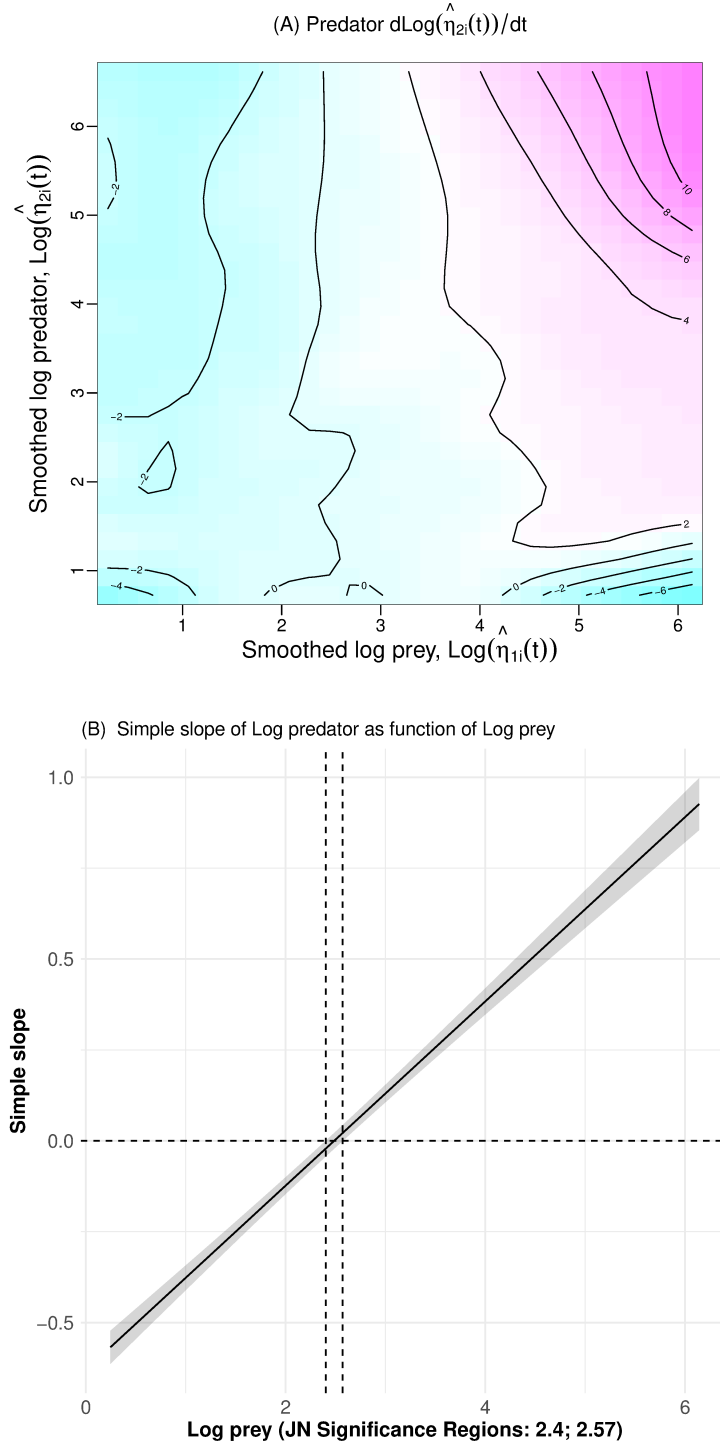


Figure 3. (A) A contour plot of the predicted $d\hat{\eta}_{1i}(t)/dt$ generated using predicted values from a GAM in which $d\hat{\eta}_{1i}(t)/dt$ was predicted using linear parametric effects of $\hat{\eta}_{1i}(t)$ and $\hat{\eta}_{2i}(t)$, as well as the tensor product between the two; (B) a plot of the simple slope of the smoothed predator density level, $\hat{\eta}_{2i}(t)$, on the prey population's smoothed first derivatives, $d\hat{\eta}_{1i}(t)/dt$, at different values of smoothed prey density, $\hat{\eta}_{1i}(t)$.

control variable yield discontinuous shifts in behavior from walking to running. The point of transition at which the shifts in behavior occur (e.g., the value of treadmill speed at which an individual switches from walking to running and vice versa) is called a bifurcation point. The goal of this illustrative example is to highlight more concretely how to use a GAM to probe for evidence of nonlinearity that contributes to bifurcation.

For this illustrative example, I focus specifically on one kind of bifurcation that may be especially applicable to the study of human dynamics: supercritical bifurcation (Strogatz, 1994). Supercritical bifurcation occurs when a stable fixed point that exists at some values of a control parameter suddenly splits into two sets of stable fixed points with continuous changes in the value of that parameter. The normal (basic or simplest) mathematical function for a supercritical bifurcation is

$$\frac{d\eta_i(t)}{dt} = r\eta_i(t) - \eta_i(t)^3, \quad (12)$$

where $\eta_i(t)$ is the true process of interest, and r is a control parameter that drives bifurcation in the system. For this system, the fixed points of the system, at which $\frac{d\eta_i(t)}{dt} = 0$, occur at three sets of possible values: 0, $\sqrt{r}$, and $-\sqrt{r}$. This is highlighted in the vector field plot in Figure 4(A)–(C). In the plot, the model-implied $\frac{d\eta_i(t)}{dt}$ values at different values of $y(t)$ are shown. Added to the vector field plot are the fixed points, shown as the intersection points between the horizontal line of $\frac{d\eta_i(t)}{dt} = 0$ and the cubic-shaped model-implied $\frac{d\eta_i(t)}{dt}$ curves. Here, the three sets of fixed points of the system are located at $\eta^* = 0$ and $\eta^* = \pm\sqrt{r}$.

Figure 4 shows that at negative values of r (see plot (A)), the fixed point at 0 is the only stable fixed point. At $r = 0$, the cubic curve becomes flatter at the origin. When $r > 0$, two new sets of stable fixed points now exist and take on the values of $\pm\sqrt{r}$. Thus, if r is varied continuously from negative to positive values, qualitative differences in the system's dynamics would arise. In particular, the number and location(s) of the fixed points would change abruptly with continuous changes in the value of r . The point $\eta^* = 0$ remains a bifurcation point whose stability cannot be determined (i.e., it is neither stable nor unstable). That is, when $r = 0$, whether a system stays at $\eta^* = 0$ depends on the system's initial values. Typically, only trajectories that start off close enough to 0 would settle into this fixed point.

A bifurcation diagram that showcases the number and values of fixed points at different values of the bifurcation parameter, r , is shown in Figure 4(D). Here, solid lines indicate the values of stable fixed points whereas the dashed line marks the fixed point at 0 that becomes unstable as r becomes greater than 0. The curve is disconnected at $r = 0$ and $\eta^* = 0$ because at $r = 0$, $\eta^* = 0$ is a bifurcation point whose stability cannot be determined. Across different values of r , it can be seen that the values of fixed points in the bifurcation diagram constitute the pattern of a pitchfork — thus giving the name of supercritical pitchfork bifurcation to this type of bifurcation.

I generated over-time trajectories using a constant time interval of 0.01 across $T = 300$ measurement occasions and 30 hypothetical participants using the ODE in Equation (12), and added normally distributed measurement errors with zero mean and variance of 1.0. Each hypothetical participant was assumed to cycle through three possible values of r : -2, 0, and 2. The over-time trajectories of 20 randomly selected subjects' true $\eta_i(t)$ (i.e., not contaminated with measurement errors), as grouped by the participants' values of r ,

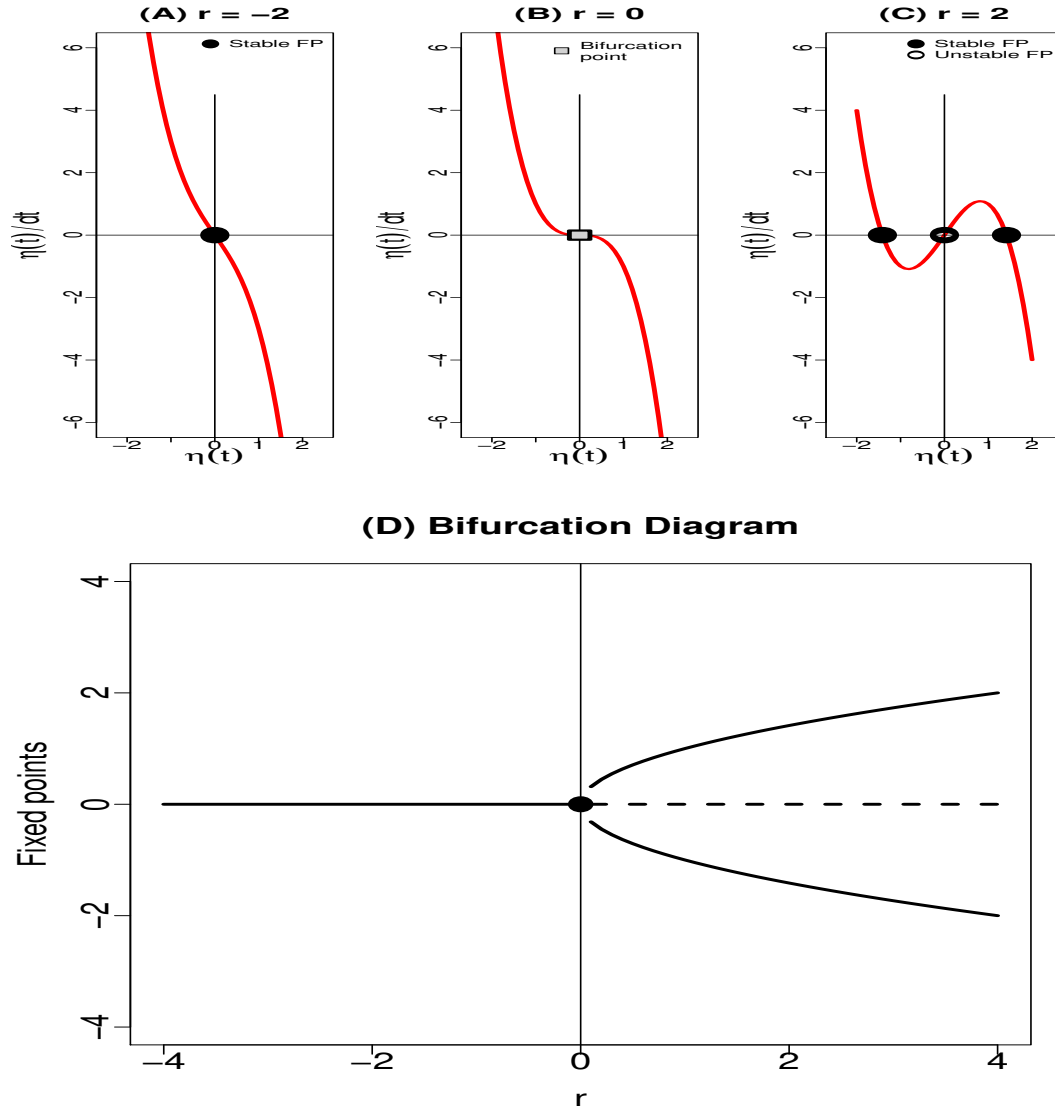


Figure 4. (A)–(C) Vector field plots showing the rates of change of a system showing supercritical bifurcation; (D) the corresponding bifurcation diagram of the system. Bifurcation point, stable and unstable fixed points are marked with gray-filled, solid, and unfilled circle, respectively, in plots (A)–(C). r is a control parameter that drives bifurcation in the system in Equation (12).

are plotted in Figure 5(A). The corresponding component-plus-residual plot of $d\hat{\eta}_i(t)/dt$ against $\hat{\eta}_i(t)$ after the linear effects of r and $\hat{\eta}_i(t)$ have been partialled out is shown in Figure 5(B). Two systematic patterns may be gleaned from the component-plus-residual plot. First, there appear to be clusters of cubic trends in the plot. Second, multiple values of $\hat{\eta}_i(t)$ yield $d\hat{\eta}_i(t)/dt$ values that are 0. Recall that values of y that give rise to a 0 rate of change in y are potential fixed points in the system. The simultaneous existence of multiple fixed points provides some initial insight into the potential nonlinearity of the system. The cubic trends further suggest the need to incorporate a cubic term involving $\hat{\eta}_i(t)$ into the fitted model. In addition, the clustering of the points in the component-plus-residual plot based on values that correspond closely to values of r also suggests the need to probe for possible interaction effect between r and $\hat{\eta}_i(t)$ on $d\hat{\eta}_i(t)/dt$. A GAM was then fitted to the smoothed level and derivative estimates in which I predicted $d\hat{\eta}_i(t)/dt$ using an intercept, a parametric interaction effect between r and $\hat{\eta}_i(t)$, and a nonparametric smooth of the effect of $\hat{\eta}_i(t)$. The estimated nonparametric smooth term for the effect of the effect of $\hat{\eta}_i(t)$ is plotted in Figure 5(C). The cubic relation is saliently captured by the nonparametric smooth term. Although not shown here, other nonparametric smooths, such as a smooth of the coefficient of $\hat{\eta}_i(t)$ as dependent on the value of r , i.e., $s(r_i)\hat{\eta}_i(t)$, may also be used to clarify the interactive relation between r and y .

Bifurcation is one of the fundamental characteristics inherent to another class of well-known dynamical systems termed catastrophe systems (Thom, 1993). In particular, if an intercept term, h , is added to the right-hand side of Equation (12), then we obtain the equilibrium points or solutions for a dynamical system known as the cusp catastrophe system (Chow, Witkiewicz, Grasman, & Maisto, 2015; Strogatz, 1994). Zeeman (1976) used the cusp catastrophe system to describe a dog's abrupt shifts in behavioral response between attacking (fight) and retreating (flight) with continuous changes in rage and fear (i.e., the control variables). Bifurcation occurs in this example as continuous changes in one of the independent variables (e.g., rage) yield sudden, qualitative changes in behavior (e.g., a shift from a single mode of outcome involving moderate behavior such as growling to the coexistence of two extreme modes of behavior, namely, attacking and retreating). Other examples include applications of the cusp catastrophe model to represent the dynamics of human driving speed (Poston & Stewart, 1978), attitude (Flay, 1978), affective states (Strahan & Conger, 1999), alcohol use (Clair, 1999; Witkiewicz, van der Maas, Hufford, & Marlatt, 2007), and developmental discontinuities (Freedle, 1977; van der Maas et al., 2003).

Despite catastrophe models' conceptual appeal and contributions, widespread applications of the catastrophe model and other related dynamic models that yield bifurcation have been impeded by challenges in identifying appropriate control parameters or variables that could drive bifurcation in human behaviors, and replicating instances of bifurcation in empirical studies involving human participants. Even though the current illustration does not solve all of these challenges, my hope is that it helps demonstrate the feasibility of performing general model exploration procedures that reveal potential nonlinear dependencies among variables and their derivatives, regardless of whether the system of interest shows bifurcation.

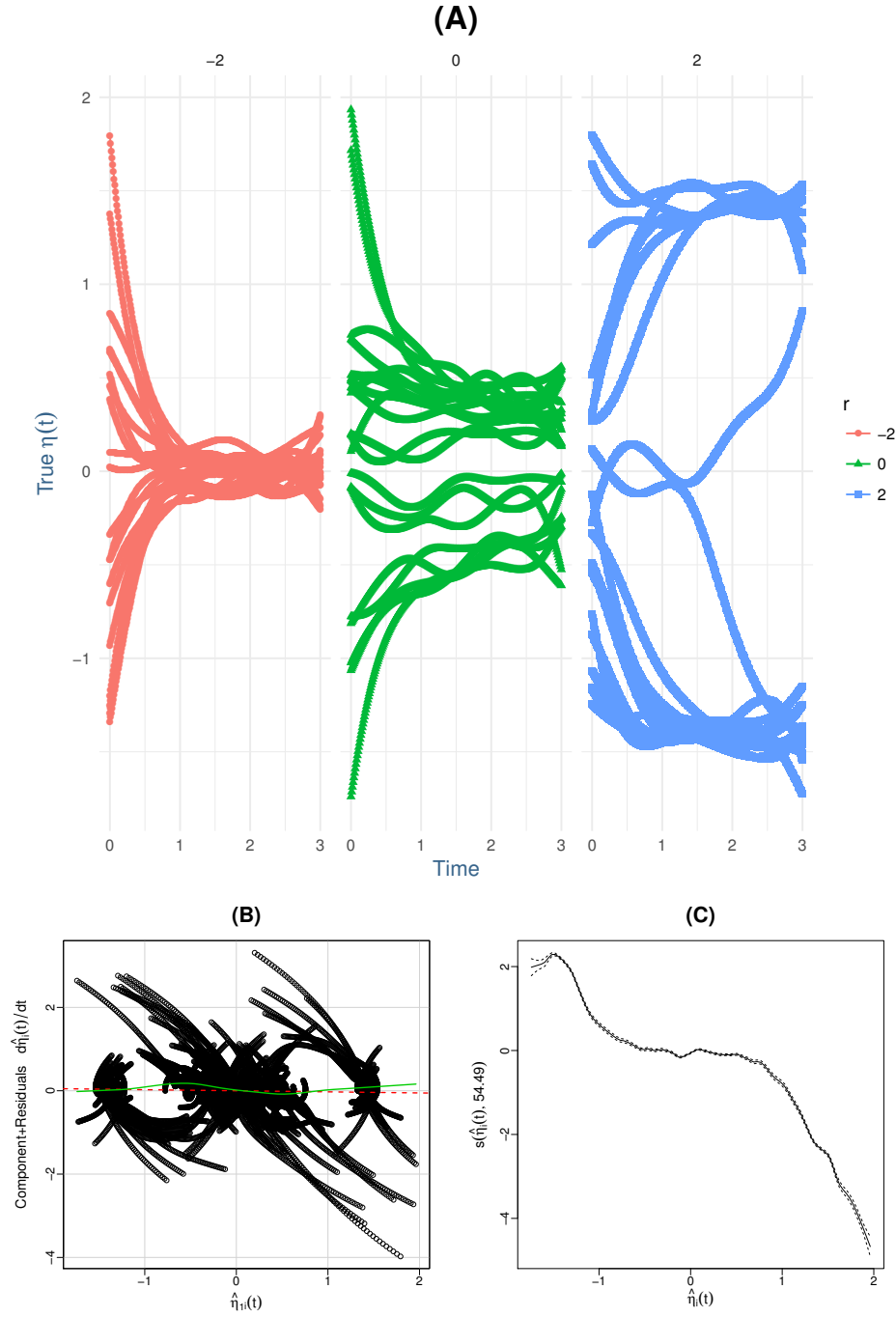


Figure 5. (A) Over-time trajectories from 20 randomly selected subjects generated three possible values for the control parameter, r , using the dynamical system model with supercritical bifurcation in Equation (12); (B) component-plus-residual plot revealing the association between $d\hat{\eta}_i(t)/dt$ and $\hat{\eta}_i(t)$ after the linear effects of r and $\hat{\eta}_i(t)$ have been partialled out; (C) estimated nonparametric smoothed effect of $\hat{\eta}_i(t)$ on $d\hat{\eta}_i(t)/dt$.

Illustration IV: Time-Varying Linear Oscillator Model

The linear oscillator model shown in Equation (9) assumes perfect constancy in how the oscillatory process unfolds over time. Now suppose the system undergoes gradual changes in the extent of damping over time. The damping parameter in this case is a time-varying parameter (TVP). My goal in this illustration is to evaluate the utility and potential limitations of the varying coefficient component of the GAM in revealing evidence of TVPs, and in regularizing (“smoothing out”) these changes to facilitate construction of parsimonious mathematical functions for confirmatory modeling purposes.

Dynamical systems with TVPs generally violate longitudinal invariance and hence the stationarity as well as stability assumptions. At first sight, this seems to violate one of the fundamental premises that enable researchers to study change in the first place: How can we study change if there is no strict constancy in how individuals change over time? The caveat here is that we assume the changes in the TVPs to occur at much slower time scales than those associated with other variables in the model (e.g., $\eta_i(t)$). As such, the processes of interest can at least be defined in a locally consistent way within segments of the data. This is one of the key assumptions that allow a model with TVPs to be estimated and for inference to be made. In this sense, such changes are not uncommon in “real-world” settings. The lead-lag influences between fluid and crystallized intelligence change as developmental changes unfold (Ferrer & McArdle, 2004). In a similar vein, affect researchers have suggested that the associations between positive and negative emotions may vary under low- as compared to high-activation scenarios (e.g., while reading a book versus during one’s College graduation ceremony; Chow et al., 2011). Still, such dynamic associations among constructs should, in principle, be changing slowly enough that homogeneity and constancy can be expected within shorter windows of time.

There is no shortage of TVP models in the social and behavioral sciences literature. For example, Molenaar (1994) considered one variation of a dynamic factor analysis model, a one-factor model with first-order autoregressive [AR(1)] process at the latent level. Polynomial functions of time were used to represent the dynamics of the TVPs, including the AR(1) and factor loading parameters. Similar polynomial functions were used by Oud and Jansen (2000) to allow for TVPs in the context of fitting linear SDE models within the structural equation modeling framework. In the econometric literature, Del Negro and Otrok (2008) examined a dynamic factor analysis model with time-varying factor loadings within a Bayesian framework. Other researchers (e.g., Stock & Watson, 2008) have also considered exploratory approaches aimed at identifying shifts in the factor loadings and time series parameters of dynamic factor analysis models. Other examples of popular TVP models are the local linear trend model (Harvey, 2001), time-varying autoregressive moving average (ARMA) model (Tarvainen et al., 2006; Weiss, 1985), and stochastic regression model (Pagan, 1980).

In previous studies, researchers have estimated TVPs by specifying them as additional latent variables that are governed by their own dynamic functions. Subsequently, a dynamic model, usually a nonparametric model or other model that is deemed flexible enough to capture a variety of different change trajectories, is used to approximate changes in the TVPs (e.g., Chow et al., 2011; Molenaar & Newell, 2003; Tarvainen et al., 2006). I refer to this approach as a confirmatory-based approach. More recently, Bringmann et al. (2017) used penalized regression splines with GAMs to explore and model TVPs. The same GAM

approach is used to probe the dynamics of TVPs in this example, but the extension to DE models is novel and has not been previously tested or illustrated.

Similar to the work by Chen, Chow, and Hunter (2018, in press), I specified the damping parameter, $\zeta_i(t)$, to be varying over time following an Ornstein-Uhlenbeck model (Oravec et al., 2016; Uhlenbeck, 1980), a model often used to describe processes with quick (exponential) returns to a set point. This model is expressed as:

$$\begin{aligned} d^2\eta_i(t) &= \left(\omega\eta_i(t) + \zeta_i(t) \frac{d\eta_i(t)}{dt} \right) dt^2 + dw_1(t), \\ \frac{d^2\zeta_i(t)}{dt} &= -\beta(\zeta_i(t) - \zeta_0) + dw_2(t). \end{aligned} \tag{13}$$

In Equation (13), $dw_1(t)$ and $dw_2(t)$ denote the differences in Wiener process over dt , assumed to be normally distributed with zero means and variances, $\sigma_{w1}^2 dt$ and $\sigma_{w2}^2 dt$, respectively. Of particular interest in this equation is the scenario where β is greater than 0, in which case the values of $\zeta_i(t)$ are assumed to approach the equilibrium ζ_0 at a rate controlled by β . The more positive β is, the faster the approach rate. This special case may be helpful for representing situations in which individuals show emotional outbursts that initially amplify over time, but later self-regulate to show progressively smaller magnitudes of emotional fluctuations.

Equation (13) additionally posits that there may be other unmeasured stochastic changes, termed process noises, that drive those individuals to deviate from a perfectly smooth and predictable oscillatory trajectory. For example, this process would occur given exposure to new perturbations that intensify a child's emotional outbursts despite the child's initial tendency to return to a homeostatic status. Unlike measurement errors which affect only one measurement occasion, the effects of these process noises would continue to affect the system's true underlying levels beyond just that particular time point.

Consider first a deterministic special case of Equation (13) with the process noise variances, σ_{w1}^2 and σ_{w2}^2 , both set to 0. Similar to the setting used in Illustration I, I set ω to -0.8, and set the initial conditions for $\eta_i(0)$ and $\frac{d\eta_i(0)}{dt}$ to be univariate normally distributed with means of 2.0 and 0, and variances of 1 and 1, respectively. I set the initial value of $\zeta_i(0)$ to 0.3 at $t = 0$ to 0.3 for all participants, β to 0.05, and ζ_0 to -0.2. These parameter values were selected to mirror the hypothetical scenario described earlier: Individuals manifesting initial amplification in emotional outbursts (with $\zeta_i(0) = 0.3$), followed by eventual return to a homeostatic status as $\zeta_i(t)$ settles into the small negative baseline value of $\zeta_0 = -0.2$. I generated over-time trajectories using Equation (13) for 10 participants across $T = 1200$ measurement occasions at a constant time interval of 0.1.

Smoothed, over-time trajectories of the data are plotted in Figure 6(A), with the component-plus-residual plot depicting the relations between the smoothed second derivative estimates and the smoothed first derivative estimates after partialling out the effects of the smoothed levels. In this case, a regular linear regression model indicated no significant association between the smoothed first and second derivatives, thus suggesting the lack of damping or amplification in the magnitude of $\hat{\eta}_i(t)$, despite clear visual evidence of over-time reduction in the amplitude of $\hat{\eta}_i(t)$ over time in Figure 6(A). Consonant with the results from the linear regression model, the loess line in the component-plus-residual plot also did not provide clear indication of the time-varying association between the first and second

derivative estimates, even though eyeballing the component-plus-residual plot revealed two clusters of points that highlight the possible existence of negative as well as positive slopes linking some of the smoothed first derivative estimates to the smoothed second derivative estimates around the area where $\frac{d\hat{\eta}_i(t)}{dt}$ is close to zero. It may be tempting to infer the existence of between-individual or between-class differences in dynamics based on Figure 6(B) alone. However, we know that in reality, there are absolutely no between-individual differences in dynamics at all in the true data generation mechanism in this particular simulation, except for individual differences in initial conditions.

I then fit a GAM with:

$$\frac{d^2\hat{\eta}_i(t)}{dt^2} = g_1\hat{\eta}_i(t) + s_1(\text{Time}_i(t))\frac{d\hat{\eta}_i(t)}{dt} + e_i(t), \quad (14)$$

in which $e_i(t)$ is assumed to be independent and normally distributed error. g_1 denotes the regression slope associated with the parametric, linear effect of $\hat{\eta}_i(t)$, and $s_1(\text{Time}_i(t))$ denotes the smoothed time-varying slope of $\frac{d\hat{\eta}_i(t)}{dt}$, plotted in Figure 6(C) over time.

The results indicated that the smooth term associated with the time-varying effect of $\frac{d\hat{\eta}_i(t)}{dt}$ on $\frac{d^2\hat{\eta}_i(t)}{dt^2}$ were statistically significant ($p < .0001$), with effective degrees of freedom (edfs) that deviated considerably from 1.0 (edf = 10.0). An edf value that deviates substantially from 1.0 suggests that the associated smooth term shows notable deviations from linearity. Edfs are inversely related to the smoothing parameter used in the penalized basis functions to smooth out “wiggleness” in the data. Roughly speaking, they may be viewed as weights that map the penalized smoothed coefficient associated with a predictor (e.g., $\frac{d\hat{\eta}_i(t)}{dt}$ in this example) to the unpenalized linear parametric coefficient associated with the predictor. An edf value that is close to zero implies that a particular predictor does not have substantial effect on the dependent variable whereas an edf value close to 1.0 suggests insufficient evidence for the effect of the predictor to be nonlinear (S. N. Wood, 2006).

Figure 6(C) suggests that the smoothed time-varying slope of $\frac{d\hat{\eta}_i(t)}{dt}$ provides a reasonable approximation for the true, over-time trajectory of $\zeta_i(t)$. However, some spurious nonlinearity in the form of amplifying oscillations is detected between $t = 80$ and 120. Another possible methodological issue is that the estimated edf is close to the starting value of the number of basis functions used, suggesting possible inadequacy in using 10 or fewer basis functions to approximate the smoothed time-varying slope of $\frac{d\hat{\eta}_i(t)}{dt}$. However, the estimated edf values continue to increase and closely mirror the starting value of the number of basis functions when the latter was increased to 15, 20, and even 200. As the number of basis functions increases beyond 10, even greater noisy oscillations were captured in the smooth term beyond $t = 80$. This result is an instance in which the penalized basis function is under-smoothing the curvatures in the time-varying slope of $\frac{d\hat{\eta}_i(t)}{dt}$.

Next, I simulated data using the same setting, with the exception that I allowed differences in the Wiener process for the oscillatory process to have a variance $\sigma_{w1}^2 = 2.25$, and $\sigma_{w2}^2 = .0009$. Thus, in this case, the oscillatory process is contaminated with both process and measurement noises, with the latter designed to be normally distributed with mean zero and variance $\sigma_e^2 = 1.0$. Applying FDA to these data yielded the smoothed level estimates plotted in Figure 7(A). Compared to the smoothed trajectories in Figure 6(A), the trajectories depicted in Figure 7(A) are characterized by greater between- and within-individual heterogeneity in the extent of damping or amplification over time. That is, even though

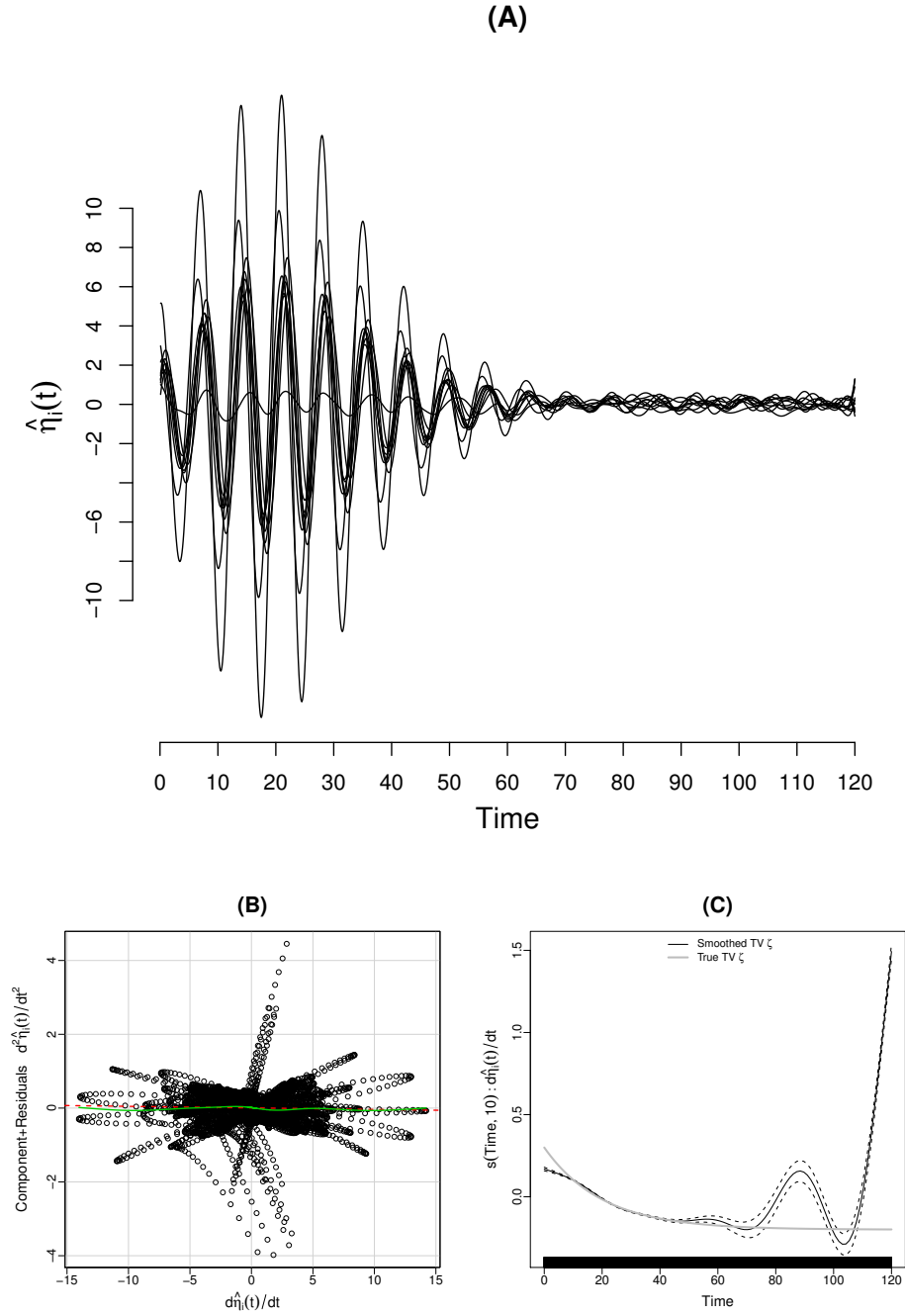


Figure 6. (A) Over-time trajectories generated using the linear oscillator model with time-varying ζ parameter; (B) component-plus-residual plot revealing the association between $d^2\hat{\eta}_i(t)/dt^2$ and $d\hat{\eta}_i(t)/dt$ after the linear effect of $\hat{\eta}_i(t)$ has been partialled out; (C) estimated smoothed, time-varying effect of $d\hat{\eta}_i(t)/dt$ on $d^2\hat{\eta}_i(t)/dt^2$ using thin-plate regression splines.

the drift function for the time-varying $\zeta_i(t)$ is identical to that used in the deterministic special case presented earlier, the addition of process noise to the data obscures some of the regularity in dynamics manifested by the individuals as a group.

As in Chen et al. (2018, in press), I used the CDEKF to fit the correctly specified stochastic time-varying oscillator model to the simulated data, with initial condition mean and variance parameters fixed at their true values, and the remaining parameters freely estimated (see supplementary materials for sample *dynr* code for fitting this model). Smoothed estimates of the time-varying $\zeta_i(t)$ and the latent oscillatory process, $\eta_i(t)$, as obtained using the CDEKF for one hypothetical subject, are shown in Figures 7 (B) and (C). The smoothed estimates of the latent variables, including the time-varying $\zeta_i(t)$ — now represented as part of the expanded latent variable vector — are generally well recovered. For instance, the 95% confidence interval (CI) for the CDEKF estimates of $\zeta_i(t)$ included the true $\zeta_i(t)$ for that subject most of the time. The parameter estimates from the CDEKF also closely approximated their true values (see Table 1).

To provide some comparisons of the results from confirmatory modeling to semiparametric results from the GAM framework, I fit the GAM in Equation (14) to the same set of simulated data. The corresponding smoothed estimates of $\zeta_i(t)$ from GAM are also plotted in Figure 7(B). Because GAM is a semiparametric representation of the original parametric model, not all of the parameters reported in Table 1 from the confirmatory framework were available or estimated. The only common parameter that was estimated in both frameworks was ω , which was estimated within the GAM framework to be -0.78 (closely mirroring the point estimate from the confirmatory model, -0.79). However, the standard error estimate based on the semiparametric GAM was 0.002 — approximately 4 times smaller than the standard error estimate from the confirmatory model (.008; see Table 1). The GAM results also returned relatively large value of estimated residual variance (1.98; no standard error estimate was provided), as compared to the estimated measurement error variance of $\hat{\sigma}_e^2$ of 0.99 from fitting the correctly specified confirmatory model.

The estimated trajectory of $\zeta_i(t)$ from the semiparametric framework in (B) represents the entire sample’s smoothed, over-time variations in $\zeta_i(t)$. This trajectory is in contrast to the estimated $\zeta_i(t)$ trajectory from the CDEKF depicted in the same plot, which shows the estimated trajectory for the one hypothetical subject whose true $\zeta_i(t)$ is also shown in the plot. The GAM estimates provided a reasonable, smoothed approximation of the true group-based drift function of $\zeta_i(t)$, even though they did capture a slight upward, spurious trend in $\zeta_i(t)$ toward the end of the data span. Note also that the 95% CI for the $\zeta(t)$ estimates from the GAM was notably narrower than that from the CDEKF (see Figure 7(C)). This is due in part to the fact that the former captures the uncertainty in the $\zeta(t)$ estimates for the entire sample, whereas the 95% CI from the CDEKF reflects the uncertainty for one hypothetical subject over time. There were some non-trivial deviations of the GAM trajectory from the group-based deterministic drift function, which predicts eventual convergence of the trajectory of $\zeta_i(t)$ toward $\zeta_0 = -0.20$. The 95% CI from GAM, unfortunately, was overly narrow and did not include the value of $\zeta_0 = -0.20$ after approximately $t_{ij} = 60$. Other parameters (e.g., β , ζ_0 , σ_{w1}^2 , σ_{w2}^2) were not available from the GAM framework and are thus not discussed here.

In summary, this final example shows that GAMs have some utility in approximating the change trajectories of TVPs in situations where insufficient theoretical knowledge exists

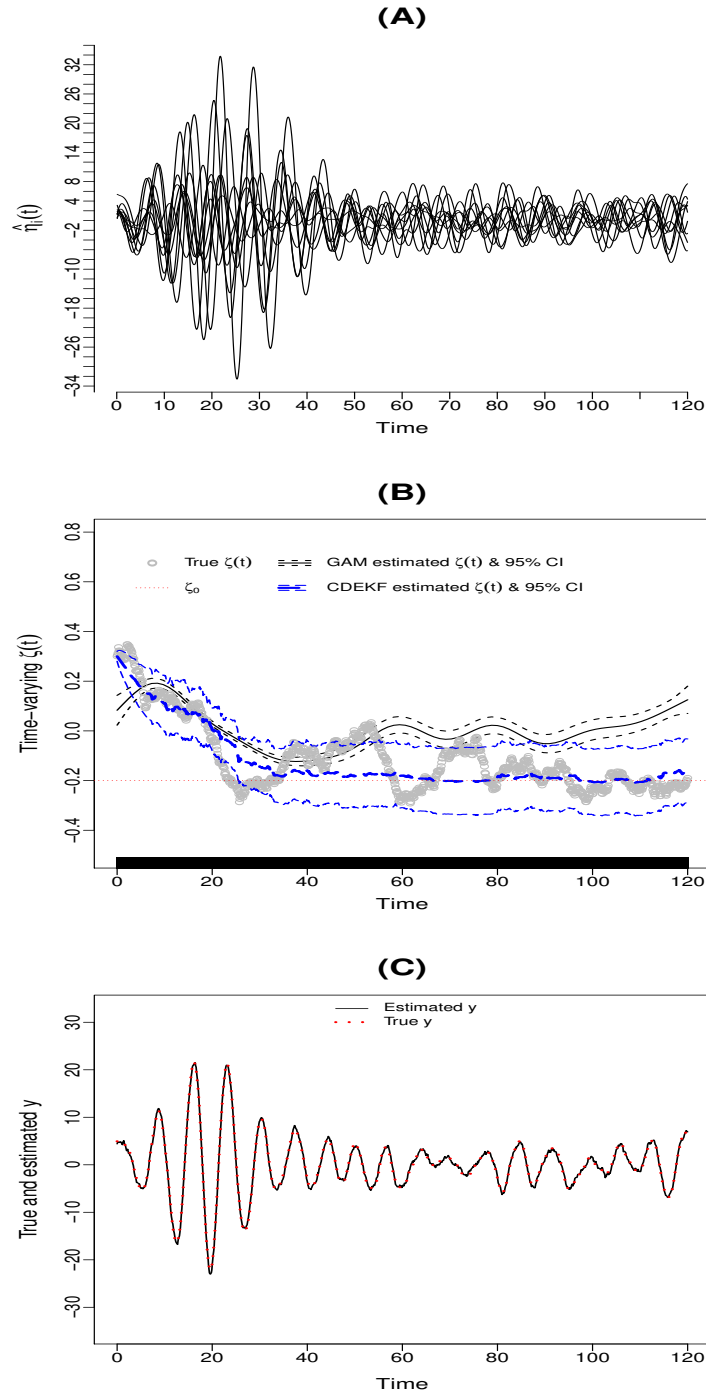


Figure 7. (A) Over-time trajectories generated using the stochastic linear oscillator model with time-varying ζ parameter; (B) the true $\zeta(t)$ trajectory (as densely overlapping circles), smooth of the time-varying weight of $\frac{d\hat{\eta}_i(t)}{dt}$ (in thin solid line) with corresponding 95% CI (in solid dashed lines) from using penalized thin-plate regression in mgcv, and smoothed estimates of $\zeta_i(t)$ (in thick long dashed line) with corresponding 95% CI (in thin long dashed lines) for one hypothetical participant using the CDEKF; (C) smoothed estimates of the latent oscillatory process, $\eta_i(t)$ (in solid line), and the true $\eta_i(t)$ values (as dots) from the same hypothetical participant using the CDEKF. CDEKF = Continuous-Discrete Extended Kalman Filter.

Table 1

True and Estimated Parameters for the Time-Varying Stochastic Oscillator Model in Illustration IV for One Replication Using the CDEKF Algorithm in dynr and GAM from mgcv.

Parameters	True values	Estimated values (standard errors) from confirmatory modeling with <i>dynr</i>	Estimated values (standard errors) from <i>mgcv</i>
ω	-0.80	-0.79 (0.008)	-0.78 (0.002)
β	0.05	0.05 (0.01)	—
ζ_0	-0.20	-0.20 (0.03)	—
σ_{w1}^2	2.25	2.29 (0.13)	—
σ_{w2}^2	0.0009	0.0005 (0.0003)	—
σ_e^2	1.00	0.99 (0.01)	1.98 (—)

dynr = the R package, *Dynamic Modeling in R*

CDEKF = Continuous-Discrete Extended Kalman Filter

mgcv = the R package, Mixed Generalized Additive Modeling (GAM) Computation Vehicle with Automatic Smoothness Estimation

to aid the construction of a confirmatory model. However, many methodological issues will remain unresolved if inferential conclusions are to be drawn based on multistage, exploratory results alone. For instance, the fitted GAMs using smoothed level and derivative estimates generally violate the independent error assumption because of the inherent within-subject time dependency in the data. In addition, the quality of the estimation (e.g., in terms of point and standard error estimates) may be compromised due to progressive accumulation of estimation errors throughout different stages of the estimation procedures. Thus, if adequate information can be garnered to construct a reasonable confirmatory dynamic model, results from confirmatory modeling should be used for inferential purposes.

Discussion

I presented four illustrative examples of using exploratory, semiparametric, and also confirmatory modeling tools to study features of dynamical systems such as nonlinear interactions, bifurcation, and time-dependent change characteristics (e.g., time-varying damping/amplification). All the illustrative examples considered in the present article involved simulated data. One immediate concern is what sorts of data in the social and behavioral sciences can support the explorations and model fitting covered in this article. The answer, in my view, is an encouraging one. The sample size configurations considered in the first three illustrations were in the range of 100 – 300 time points, and $n = 10 - 50$ participants. The last illustration was based on data with relatively large T and small n ($T = 1200$, $n = 10$) to highlight the specific nature of the TVP considered — the damping/amplification parameter, whose effects tend to emerge more slowly (requiring multiple iterations of cycles) than those associated with other parameters. Simulation studies elsewhere involving dynamic models with other TVPs have suggested reasonable results with larger n (e.g., 100 – 300), and T that mirrors those used in Illustrations I–III (e.g., $T = 50 - 300$).

The sample size configurations noted above are becoming more common. Many laboratory studies now include planned designs to obtain second-by-second coding of individuals'

behaviors over the course of laboratory tasks that typically last between 5 to 15 minutes. Such coded data have been used in the past for dynamical systems modeling (Cole, Ben-dezú, Ram, & Chow, 2017; Morales et al., 2018). In a similar vein, an increasing number of experience sampling studies now feature time-series between 50 and 800 occasions, often from 100+ participants (Ram, Shiyko, Lunkenheimer, Doerksen, & Conroy, 2014), all of which produce data that are conducive for dynamical systems modeling. At faster time-scales, studies involving ambulatory physiological, sleep and physical activity data routinely yield multiple-subject time series that span many thousands of occasions. Generally, the growing repertoire of intensive longitudinal data produced by advances in mobile and web technology, miniaturization of sensors, and widespread adoption of digital communication platforms speaks directly to the need to develop better and more powerful dynamic modeling tools.

Beyond studies with intensive longitudinal data, several researchers have also tested theories of change using panel data and longitudinal models that have a “dynamical systems flair” (i.e., in their emphasis on representing change mechanisms; see e.g., Ferrer & McArdle, 2004). It is plausible to test theories of change using data of limited time lengths from a large number of subjects, but only under some conditions (Molenaar, 2004). One such condition is that the change and measurement functions characterizing all individuals are homogeneous in nature. This assumption may not be tenable and should be relaxed as needed. Relatedly, data of limited time lengths or those measured at overly coarse intervals may lack power to distinguish different types of intraindividual changes from each other (e.g., diurnal from event-related variations in emotions), and from confounds such as interindividual differences in initial conditions.

Some researchers may be interested in adopting a mixed effects framework to model change patterns across multiple participants. Within this framework, individuals are postulated to conform to the same change mechanisms, but with some between-individual differences in selected modeling parameters. In this case, the techniques reviewed thus far support the evaluation of the fixed effects, namely, the effects that describe the population change trajectory as a whole. Selected extensions can be performed to allow for random effects for key parameters to be represented as latent variables in confirmatory model fitting (Ou, 2018), or estimated with Monte Carlo sampling techniques (see e.g., Chow, Lu, et al., 2016, and the references therein). The GAM framework as implemented in the *mgcv* package also allows for the inclusion of additional linear parametric random effects. Another alternative is to consider multiple-group and latent class extensions that allow different groups/latent classes to conform to distinct dynamics (Chow et al., 2018). If sufficient data are available from each individual and high degrees of between-person heterogeneity are expected, model exploration and fitting should be performed at the individual level.

In obtaining derivative estimates, the formulation in Equations (4) — (6) assumes that the same smoothing parameter and a group-based penalty term are used to derive approximation curves for all units (e.g., individuals). These constraints can also be relaxed as appropriate. That is, curve approximation and derivative estimation can be performed separately for each unit of analysis, including using individual-specific smoothing parameters to customize the amount of smoothing applied to each individual’s data. In situations where the data from different units of analysis follow trajectories that are out of phase relative to each other, a procedure called *curve registration* should first be performed to ensure that

the peaks and troughs in all units' curves are aligned with each other (Ramsay & Silverman, 2005).

The methods demonstrated in the present article are by no means exhaustive, nor are they without their limitations. For instance, I used one particular smoothing routine within the *mgcv* library, thin-plate spline regression splines (S. N. Wood, 2003, 2006), to automate the estimation of selected nonparametric effects. In practice, a variety of spline or penalized spline functions may be used to obtain the smoothed values [i.e., all terms involving $s(.)$ in Equation (8)]. Popular choices include cubic splines, B-splines, P-splines and other penalized regression splines (Green & Silverman, 1994). The thin plate regression splines have the advantages of: (a) not having to choose knot locations, thereby reducing subjectivity in modeling and selection of optimal basis functions (S. N. Wood, 2006); and (b) being able to accommodate a higher number of predictors than other spline regression methods. The same routines within the *mgcv* package have also been utilized by Bringmann et al. (2017) to fit discrete-time vector autoregressive models with time-varying coefficients. Despite the practical advantages of this approach, caution should still be exercised because complete reliance on model exploration indices such as edfs to explore the interrelations among multiple noisy variables can often lead to instances of under- or even over-smoothing. The final number of selected knot points and the corresponding edfs may also be sensitive to the starting values specified by the user, despite the robustness of such approaches to initial knot point specifications when used to approximate simpler nonparametric trends. In addition, given that multivariate models and problems are at the core of the modeling work in the social and behavioral sciences, more thorough investigation of the performance of such nonparametric approaches in cases involving multiple dependent variables is essential.

Even though I focused on FDA as the derivative estimation approach of choice in the present article, other approaches, such as the GOLD and the GLLA are also viable approaches for this purpose. Chow, Bendezú, et al. (2016) provided comparisons among the FDA, GOLD and the GLLA. It may be helpful to note a few distinctions here. First, the derivative estimates from the GLLA and GOLD tend to be “rougher” compared to those from the FDA due to the explicit use of a penalized regulation approach via Equations (6) — (7) in the FDA. Second, some data reduction always occurs in the cases of the GLLA and GOLD due to the particular data processing procedure used in these approaches (i.e., time delay embedding). In contrast, in the FDA, the sample size is always equal to the original available sample size regardless of the placement of knot points. Third, the FDA can be used readily with either equally spaced or irregularly spaced time series data. In the GLLA as well as GOLD, it is possible to make adaptations to account for irregularly spaced time intervals. But current implementation of GLLA and GOLD has not yet incorporated these adaptations. Finally, even though it may appear, at first glance, that use of the FDA approach requires quite a few decisions on multiple fronts (perhaps more so than the GLLA and GOLD), most of these decisions can be automated due to the availability of well-established guidelines. The remaining decisions can be made in relatively straightforward ways using output from freely available software packages.

The CDEKF and related algorithms implemented in the *dynr* package are but one possible way of fitting confirmatory DE models to Gaussian distributed data. Other alternative software packages include the Continuous Time Structural Equation Modeling (*ctsem*; Driver et al., 2017), *dlim* (Petrus, 2010), *KFAS* (Helske, 2017), *dse* (Gilbert, 2015),

OpenMx (Neale et al., 2016) and *bssm* (Helske & Vihola, 2018) in R, and the Bivariate Hierarchical -Uhlenbeck Model toolbox in Matlab (*BHOUM*; Oravecz et al., 2016). Unlike the *dynr* package, *ctsem*, *dlim*, *KFAS*, *dse*, *OpenMx*, and *BHOUM* only allow for linear dynamic models, but they have other unique strengths in fitting particular types of linear models. For a review, see Petris and Petrone (2011) and Ou et al. (2018, revised and resubmitted).

I have limited the scope of the present article to Gaussian distributed, continuous observed data. Categorical and other non-Gaussian data are quickly becoming the norm in many empirical studies. There is thus a clear need to extend the approaches considered in this article (e.g., the derivative estimation approaches, the use of GAMing for model exploration purposes) to accommodate such data. Alternative model exploration and estimation approaches amenable to non-Gaussian data (e.g., Bayesian and simulation-based approaches; Doucet, de Freitas, & Gordon, 2001; Durbin & Koopman, 2001; M. West & Harrison, 1997) and associated software packages (e.g., the *bssm* package; Helske & Vihola, 2018) are available.

Model identification in DE models is another key issue that warrants more attention from researchers. One necessary condition for DE models to be empirically identified is for them to be *observable*, or “estimable” from available observed data. When a system is observable, then the values of the system at any time can be fully and uniquely determined from observed measurements over a finite time interval (see p. 28; Bar-Shalom et al., 2001). Ou (2018) considered an approach in fitting nonlinear stochastic DE models in which the random effects are included and estimated as part of the latent variable vector, similar to how TVPs are included as part of the latent variable vector in Illustrative Example IV. One important result from Ou (2018) was that such a system is only observable if the number of random effects is no more than p (i.e., the number of manifest variables) or w (i.e., the number of latent variables), whichever one is smaller. Similar steps can be applied in the present context to prove that the system with TVPs is only observable if the total number of new latent variables introduced by the inclusion of the TVPs is no more than p or w , whichever one is smaller. A related issue with respect to identification is controllability, which ensures that the values of the latent variables can be controlled through manipulation of elements in the vector of exogenous variables, $\mathbf{u}_i(t)$. For the specific applications considered in the present article, the issue of controllability does not arise.

In fitting dynamical systems models, it is often strategic to incorporate theoretically driven constraints to ensure that the fitted model yields sensible values. Such constraints not only aid estimation, but are sometimes critical for model identification. For the models considered in the present article, it is possible to apply appropriate transformation functions on selected parameters to yield transformed parameters that are unconstrained during the optimization process to facilitate model estimation. For instance, optimization of parameters that are expected to be positive, such as all the parameters in ODE of the classical predator-prey model, can be performed on a log scale so these parameters are unconstrained during the optimization process. Still, in some limited cases, some of the parameters may be located on the boundary of the parameter space over which optimization is done, thereby violating regularity conditions for performing standard model comparison practices such as likelihood ratio tests (Savalei & Kolenikov, 2008).

More generally, most contemporary model fit indices designed to inform absolute fit of

latent variable models (for a review see e.g., Hu & Bentler, 1998; McDonald & Ho, 2002; S. G. West, Taylor, & Wu, 2012) may not be as useful for assessing the fit of dynamical system models. This is because such indices of absolute fit do not help pinpoint how, in what ways, and for which time points a hypothesized dynamical systems model fails to approximate key summary statistics (e.g., means and variance-covariance structures) of the data. Information criterion measures such as the Akaike Information Criterion (Akaike, 1973) and the Bayesian Information Criterion (Schwarz, 1978) have been used and found to yield reasonable model selection results in some special cases (Chow et al., 2015). Still, much remains to be done.

Along a similar line, extension of regression diagnostic analyses to the context of DE models is another important topic that has received limited attention in the literature. In particular, it is possible to develop diagnostics similar to those considered elsewhere (Chow, Hamaker, & Allaire, 2009) to identify outlying cases (both in terms of individuals and measurement occasions), and modeling features that are influential to modeling results. The need for formal diagnostic analysis and corresponding tools is especially critical in the context of dynamical systems modeling given that some nonlinear or more complex patterns of change could either arise as part of the true underlying dynamics of a system, or as driven by outliers. Finally, tools for diagnosing optimal sampling intervals to better facilitate reconstruction of the underlying dynamics of a system are also crucially needed.

Closing Remarks

Due in part to concerted efforts by groups of researchers to provide accessible tools for studying dynamical systems (e.g., Boker & Graham, 1998; Deboeck, 2010; Molenaar & Newell, 2003), the last decades have seen a steady growth in enthusiasm for dynamical systems research in the social and behavioral sciences. Applications of DE models, particularly nonlinear DEs, require mastery of some technical knowledge that may not be accessible to all social and behavioral scientists (Kolata, 1977). At times, mathematical concepts may be misconstrued or misrepresented, much as how the analogy of a butterfly flapping its wings — famously used by Lorenz as an exaggeration of the property of sensitive dependence on initial conditions manifested by the weather system — has been distorted by popular media. With few exceptions (Smith & Thelen, 1993; Vallacher & Nowak, 1994), the scarcity of working theoretical knowledge on the ways in which a system may manifest changes has deterred many researchers from undertaking a dynamic perspective to formulating research questions. Indeed, many hurdles still exist before we can broadly integrate dynamical systems concepts and ideas into mainstream research. My intent in this article is to remove some of those hurdles by presenting and illustrating some possible exploratory and confirmatory techniques as possible first steps for discovering, describing, and understanding linear as well as nonlinear human dynamics.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage. doi: 10.1037/0021-9010.84.6.897
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory*, (p. 267-281). Budapest, Hungary: Akademiai Kiado. doi: 10.1007/978-1-4612-1694-0_15
- Arminger, G. (1986). Linear stochastic differential equation models for panel data with unobserved variables. In N. Tuma (Ed.), *Sociological methodology* (p. 187-212). San Francisco, CA: Jossey-Bass. doi: 10.1080/0022250X.2010.532259
- Arnold, L. (1974). *Stochastic differential equations*. New York, NY: Wiley. doi: 10.1002/zamm.19770570413
- Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation: Theory algorithms and software*. New York, NY: Wiley. doi: 10.1002/0471221279
- Beskos, A., Papaspiliopoulos, O., & Roberts, G. (2009). Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *The Annals of Statistics*, 37(1), 223-245. doi: 10.1214/07-AOS550
- Boker, S. M., Deboeck, P. R., Edler, C., & Keel, P. K. (2010). Generalized local linear approximation of derivatives from time series. In S.-M. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (p. 161-178). New York, NY: Taylor & Francis. doi: 10.4324/9780203864746
- Boker, S. M., & Graham, J. (1998). A dynamical systems analysis of adolescent substance abuse. *Multivariate Behavioral Research*, 33, 479-507. doi: 10.1207/s15327906mbr3304_3
- Boker, S. M., & McArdle, J. J. (1995). Statistical vector field analysis applied to mixed cross-sectional and longitudinal data. *Experimental Aging Research*, 21, 77-93. doi: 10.1080/03610739508254269
- Boker, S. M., Neale, M. C., & Klump, K. L. (2014). A differential equations model for the ovarian hormone cycle. In P. C. M. Molenaar, K. Newell, & R. Lerner (Eds.), *Handbook of relational developmental systems: Emerging methods and concepts* (p. 369-391). New York, NY, USA: Guilford Press.
- Boker, S. M., & Nesselroade, J. R. (2002). A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave panel data. *Multivariate Behavioral Research*, 37, 127-160.
- Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized modeling. *Psychological Methods*, 22, 409-425. doi: 10.1037/met0000085
- Brown, E. N., & Luithardt, H. (1999). Statistical model building and model criticism for human circadian data. *Journal of Biological Rhythms*, 14, 609-616. doi: 10.1177/074873099129000975
- Brown, R. (1828). A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical*

- Magazine*, 4, 161–173. doi: 10.1017/CBO9781107775473.016
- Butner, J. E., Gagnon, K. T., Geuss, M. N., Lessard, D. A., & Story, T. N. (2015). Utilizing topology to generate and test theories of change. *Psychological Methods*, 20(1), 1-25. doi: 10.1037/a0037802
- Chen, M., Chow, S.-M., & Hunter, M. (2018, in press). Stochastic differential equation models with time-varying parameters. In K. van Montford, H. Oud, & M. Voelkle (Eds.), *Continuous-time modeling in the behavioral and related sciences*. Berlin, Germany: Springer-Verlag.
- Chow, S.-M., Bendezú, J. J., Cole, P. M., & Ram, N. (2016). A comparison of two-stage approaches for fitting nonlinear ordinary differential equation (ODE) models with mixed effects. *Multivariate Behavioral Research*, 51(2-3), 154-184. doi: 10.1080/00273171.2015.1123138
- Chow, S.-M., Ferrer, E., & Nesselroade, J. R. (2007). An unscented Kalman filter approach to the estimation of nonlinear dynamical systems models. *Multivariate Behavioral Research*, 42(2), 283-321. doi: 10.1080/00273170701360423
- Chow, S.-M., Hamaker, E. J., & Allaire, J. C. (2009). Using innovative outliers to detecting discrete shifts in dynamics in group-based state-space models. *Multivariate Behavioral Research*, 44, 465-496. doi: 10.1080/00273170903103324
- Chow, S.-M., Ho, M.-H. R., Hamaker, E. J., & Dolan, C. V. (2010). Equivalences and differences between structural equation and state-space modeling frameworks. *Structural Equation Modeling*, 17(303-332). doi: 10.1080/10705511003661553
- Chow, S.-M., Lu, Z., Sherwood, A., & Zhu, H. (2016). Fitting nonlinear ordinary differential equation models with random effects and unknown initial conditions using the Stochastic Approximation Expectation Maximization (SAEM) algorithm. *Psychometrika*, 81, 102-134. doi: 10.1007/s11336-014-9431-z
- Chow, S.-M., & Nesselroade, J. R. (2004). General slowing or decreased inhibition? Mathematical models of age differences in cognitive functioning. *Journals of Gerontology Series B: Psychological Sciences & Social Sciences*, 59B(3), 101-109. doi: 10.1093/geronb/59.3.P101
- Chow, S.-M., Ou, L., Ciptadi, A., Prince, E., Hunter, M. D., You, D., ... Messinger, D. S. (2018). Representing sudden shifts in intensive dyadic interaction data using differential equation models with regime switching. *Psychometrika*, 83(2), 476–510. doi: 10.1007/s11336-018-9605-1
- Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as thermostat: Representing emotion regulation using a damped oscillator model. *Emotion*, 5(2), 208-225. doi: 10.1037/1528-3542.5.2.208
- Chow, S.-M., Witkiewitz, K., Grasman, R. P. P. P., & Maisto, S. A. (2015). The cusp catastrophe model as cross-sectional and longitudinal mixture structural equation models. *Psychological Methods*, 20, 142-164. doi: 10.1037/a0038962
- Chow, S.-M., Zu, J., Shifren, K., & Zhang, G. (2011). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Research*, 46(2), 303-339. doi: 10.1080/00273171.2011.563697
- Clair, S. (1999). A cusp catastrophe model for adolescent alcohol use: An empirical test. *Nonlinear Dynamics, Psychology, and Life Sciences*, 2(3), 217-241. doi: 10.1023/A:102237600

- Cobb, L., & Zacks, S. (1985). Applications of catastrophe theory for statistical modelling in the biosciences. *Journal of the American Statistical Association*, 80(392), 793-802. doi: 10.1080/01621459.1985.10478184
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (n.d.). (3rd ed.). Mahway, NJ: Lawrence Erlbaum. doi: 10.4324/9780203774441
- Cole, P. M., Bendezú, J. J., Ram, N., & Chow, S.-M. (2017). Dynamical systems modeling of early childhood self-regulation. *Emotion*, 17, 684-699. doi: 10.1037/emo0000268
- Coleman, J. S. (1968). The mathematical study of change. In H. M. Blalock, Jr. & A. Blalock (Eds.), *Methodology in social research* (p. 428-478). New York, NY: McGraw-Hill.
- Craven, P., & Wahba, G. (1978, Dec 01). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4), 377-403. Retrieved from <https://doi.org/10.1007/BF01404567> doi: 10.1007/BF01404567
- De Boor, C. (1977). A package for calculating with B-splines. *SIAM Journal of Numerical Analysis*, 14, 441-472. doi: 10.1137/0714026
- De Boor, C. (1978). *A practical guide to splines*. New York, NY: Springer-Verlag. doi: 10.1002/zamm.19800600129
- Deboeck, P. R. (2010). Estimating dynamical systems: Derivative estimation hints from Sir Ronald A. Fisher. *Multivariate Behavioral Research*, 45, 725-745. doi: 10.1080/00273171.2010.498294
- Deboeck, P. R., Montpetit, M. A., Bergeman, C. S., & Boker, S. M. (2009). Using derivative estimates to describe intraindividual variability at multiple time scales. *Psychological Methods*, 14(4), 367-386. doi: 10.1037/a0016622
- Del Negro, M., & Otrok, C. (2008). *Dynamic factor models with time-varying parameters: Measuring changes in international business cycles* (Staff Reports No. 326). New York, NY: Federal Reserve Bank of New York, NY. doi: 10.2139/ssrn.1136163
- Dierckx, P. (1993). *Curve and surface fitting with splines*. New York, NY, USA: Oxford Science Publications.
- Dolan, C. V., & Molenaar, P. C. M. (1991). A note on the calculation of latent trajectories in the quasi Markov simplex model by means of regression method and the discrete Kalman filter. *Kwantitatieve Methoden*, 38, 29-44.
- Doucet, A., de Freitas, N., & Gordon, N. (Eds.). (2001). *Sequential Monte Carlo methods in practice*. New York, NY: Springer. doi: 10.1007/978-1-4757-3437-9
- Driver, C., Oud, J., & Voelkle, M. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software, Articles*, 77(5), 1-35. Retrieved from <https://www.jstatsoft.org/v077/i05> doi: 10.18637/jss.v077.i05
- Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199641178.001.0001
- Eilers, P., & Marx, B. (1996). Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). *Statistical Science*, 11(2), 89-121. doi: 10.1214/ss/1038425655
- Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology*, 40(6), 935-952. doi: 10.1037/0012-1649.40.6.935
- Flay, B. R. (1978). Catastrophe theory in social psychology: Some applications to attitudes and social behaviors. *Behavioral Science*, 23(335-350). doi: 10.1002/bs.3830230404

- Fox, J. (2015). *Applied regression analysis and generalized linear models* (3rd ed.). New York, NY: Sage.
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks, CA: Sage. Retrieved from <http://tinyurl.com/carbook>
- Freedle, R. (1977). Psychology, Thomian topologies, deviant logics, and human development. In N. Datan & H. W. Reese (Eds.), *Life-span developmental psychology: Dialectical perspectives on experimental research* (p. 317-342). New York, NY: Academic Press.
- Geweke, J. (1996). Variable selection and model comparison in regression. In J. O. Berger, J. M. Bernardo, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics* (5th ed., p. 609-620). Oxford, United Kingdom: Oxford University Press.
- Gilbert, P. D. (2015). Brief user's guide: Dynamic systems estimation [Computer software manual]. Retrieved from <http://cran.r-project.org/web/packages/dse/vignettes/Guide.pdf>
- Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Boca Raton, FL: Chapman & Hall/CRC. doi: 10.1007/978-1-4899-4473-3
- Harrell, F. E. J. (2001). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*. New York, NY: Springer. doi: 10.1007/978-1-4757-3462-1
- Harvey, A. C. (2001). *Forecasting, structural time series models and the Kalman filter*. Cambridge, United Kingdom: Cambridge University Press.
- Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. London, United Kingdom: Chapman & Hall. doi: 10.1201/9780203753781
- Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society series B (Methodological)*, 55(4), 757-796.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction (2nd)*. Hoboken, New Jersey.: Springer.
- Hawkey, L. C., Burleson, M. H., Berntson, G. G., & Cacioppo, J. T. (2003). Loneliness in everyday life: Cardiovascular activity, psychosocial context and health behaviors. *Journal of Personality and Social Psychology: Personality and Individual Differences*, 85(1), 105-120. doi: 10.1037/0022-3514.85.1.105
- Helske, J. (2017). KFAS: Exponential family state space models in R. *Journal of Statistical Software, Articles*, 78(10), 1-39. doi: 10.18637/jss.v078.i10
- Helske, J., & Vihola, M. (2018). bssm: Bayesian inference of non-linear and non-Gaussian state space models in R [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=bssm/vignettes/bssm.pdf> (R package version 0.1.5)
- Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424. doi: 10.1037//1082-989x.3.4.424
- Kaplan, D., & Glass, L. (1995). *Understanding nonlinear dynamics*. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4612-0823-5
- Kolata, G. (1977). Catastrophe theory: The emperor has no clothes. *Science*, 196, 287-351. doi: 10.1126/science.196.4287.287
- Kou, S. C., Olding, B. P., Lysy, M., & Liu, J. S. (2012). A multiresolution method

- for parameter estimation of diffusion processes. *Journal of the American Statistical Association*, 107(500), 1558-1574. Retrieved from <https://doi.org/10.1080/01621459.2012.720899> (PMID: 25328259) doi: 10.1080/01621459.2012.720899
- Kulikov, G., & Kulikova, M. (2014, Jan). Accurate numerical implementation of the continuous-discrete extended Kalman filter. *Automatic Control, IEEE Transactions on*, 59(1), 273-279. doi: 10.1109/TAC.2013.2272136
- Kulikova, M. V., & Kulikov, G. Y. (2014). Adaptive ODE solvers in extended Kalman filtering algorithms. *Journal of Computational and Applied Mathematics*, 262, 205-216. doi: 10.1016/j.cam.2013.09.064
- Lawley, D. N., & Maxwell, M. A. (1973). Regression and factor analysis. *Biometrika*, 60(2), 331-338. Retrieved from <https://doi.org/10.1093/biomet/60.2.331> doi: 10.1093/biomet/60.2.331
- Li, R., Tan, X., Huang, L., Wagner, A. T., & Yang, J. (2014). TVEM (time-varying effect model) SAS macro suite users' guide (version 2.1.1) [Computer software manual]. Retrieved from <http://methodology.psu.edu>
- Lotka, A. J. (1925). *Elements of physical biology*. Baltimore: Williams and Wilkins.
- Lu, Z.-H., Chow, S.-M., & Loken, E. (2017). A comparison of Bayesian and frequentist model selection methods for factor analysis models. *Psychological Methods*, 22, 361-381. doi: 10.1037/met0000145
- Lu, Z.-H., Chow, S.-M., Sherwood, A., & Zhu, H. (2015). Bayesian analysis of ambulatory cardiovascular dynamics with application to irregularly spaced sparse data. *Annals of Applied Statistics*, 9, 1601-1620. doi: 10.1214/15-AOAS846
- Mbalawata, I. S., Särkkä, S., & Haario, H. (2013). Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and non-linear Kalman filtering. *Computational Statistics*, 28(3), 1195-1223. doi: 10.1007/s00180-012-0352-y
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London, United Kingdom: Chapman and Hall. doi: 10.1007/978-1-4899-3242-6
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7, 64-82. doi: 10.1037/1082-989X.7.1.64
- McKeown, G. J., & Sneddon, I. (2014). Modeling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychological Methods*, 19, 155-174. doi: 10.1037/a0034282
- Merrilees, C. E., Goeke-Morey, M., & Cummings, E. M. (2008). Do event-contingent diaries about marital conflict change marital interactions? *Behaviour Research and Therapy*, 46, 253-262. doi: 10.1016/j.brat.2007.11.007
- Molenaar, P. C. M. (1994). Dynamic latent variable models in developmental psychology. In A. von Eye & C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (p. 155-180). Thousand Oaks, CA: Sage Publications.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology — this time forever. *Measurement: Interdisciplinary Research and Perspectives*, 2, 201-218. doi: 10.1207/s15366359mea0204_1
- Molenaar, P. C. M., Beltz, A. M., Gates, K. M., & Wilson, S. J. (2007). State space modeling of time-varying contemporaneous and lagged relations in connectivity maps. *NeuroImage*, 125, 791-802. Retrieved from <http://dx.doi.org/10.1016/j.neuroimage.2015.10.088>

- Molenaar, P. C. M., & Newell, K. M. (2003). Direct fit of a theoretical model of phase transition in oscillatory finger motions. *British Journal of Mathematical and Statistical Psychology*, 56, 199-214. doi: 10.1348/000711003770480002
- Morales, S., Ram, N., Buss, K. A., Cole, P. M., Helm, J. L., & Chow, S.-M. (2018). Age-related changes in the dynamics of fear-related regulation in early childhood. *Developmental Science*, 21(5), e12633. doi: 10.1111/desc.12633
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., ... Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 80(2), 535-549. doi: 10.1007/s11336-014-9435-8
- Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2016). Bayesian data analysis with the bivariate hierarchical Ornstein-Uhlenbeck process model. *Multivariate Behavioral Research*, 51, 106-119. doi: 10.1080/00273171.2015.1110512
- Ou, L. (2018). *Estimation of mixed effects continuous-time models* (Unpublished doctoral dissertation). Pennsylvania State University. (Ph.D. dissertation)
- Ou, L., Hunter, M., & Chow, S.-M. (2018, revised and resubmitted). What's for dynr: A package for linear and nonlinear dynamic modeling in R. *The R Journal*.
- Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, 65(2), 199-215. doi: 10.1007/bf02294374
- Oud, J. H. L., van den Bercken, J. H., & Essers, R. J. (1990). Longitudinal factor score estimation using the Kalman filter. *Applied Psychological Measurement*, 14, 395-418. doi: DOI:10.1177/014662169001400406
- Pagan, A. (1980). Some identification and estimation results for regression models with stochastically varying coefficients. *Journal of Econometrics*, 13, 341-363. doi: 10.1016/0304-4076(80)90084-6
- Petris, G. (2010). An R package for dynamic linear models. *Journal of Statistical Software*, 36(12), 1-16. doi: 10.18637/jss.v036.i12
- Petris, G., & Petrone, S. (2011). State space models in R. *Journal of Statistical Software*, 41(4), 1-25. doi: 10.18637/jss.v041.i04
- Poston, T., & Stewart, I. N. (1978). *Catastrophe theory and its applications*. London, United Kingdom: Pitman.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical recipes in C*. Cambridge, United Kingdom: Cambridge University Press.
- Ram, N., Shiyko, M., Lunkenheimer, E. S., Doerksen, S., & Conroy, D. (2014). Families as coordinated symbiotic systems: Making use of nonlinear dynamic models. In A. In Booth, S. McHale, & N. Landale (Eds.), *Emerging methods in family research, 4th national symposium on family issues* (p. 19-37). Cham, Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-01562-0_2
- Ramsay, J. O., Hooker, G., Campbell, D., & Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *Journal of Royal Statistical Society: Series B*, 69(5), 741-796. doi: 10.1111/j.1467-9868.2007.00610.x
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). Springer-Verlag: New York, NY. doi: 10.1007/b98888
- Savalei, V., & Kolenikov, S. (2008). Constrained versus unconstrained estimation in structural equation modeling. *Psychological Methods*, 13(2), 150-170. doi: 10.1037/

1082-989x.13.2.150

- Scheinerman, E. R. (1996). *Invitation to dynamical systems*. New Jersey: Prentice Hall.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464. doi: 10.1214/aos/1176344136
- Singer, H. (2002). Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood vs. extended Kalman filter and Itô-Taylor expansion. *Journal of Computational and Graphical Statistics*, 11, 972-995. doi: 10.1198/106186002808
- Singer, H. (2010). SEM modeling with singular moment matrices part I: ML-estimation of time series. *The Journal of Mathematical Sociology*, 34(4), 301-320. doi: 10.1080/0022250X.2010.532259
- Singer, H. (2012). SEM modeling with singular moment matrices part II: ML-estimation of sampled stochastic differential equations. *The Journal of Mathematical Sociology*, 36(1), 22-43. doi: 10.1080/0022250X.2010.532259
- Smith, L. B., & Thelen, E. (1993). *A dynamic systems approach to development*. Cambridge, MA: MIT Press.
- Stock, J., & Watson, M. (2008). Forecasting in dynamic factor models subject to structural instability. In J. Castle & N. Shephard (Eds.), *The methodology and practice of econometrics, a Festschrift in honour of Professor David F. Hendry*. Oxford, United Kingdom: Oxford University Press.
- Strahan, E. Y., & Conger, A. J. (1999). Social anxiety and social performance: Why don't we see more catastrophes. *Journal of Anxiety Disorders*, 13(4), 399-416. doi: 10.1016/s0887-6185(99)00006-7
- Strogatz, S. H. (1994). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Cambridge, MA: Westview. doi: 10.1201/9780429492563
- Tan, X., Shiyko, M. P., Li, R., Li, Y., & Dierker, L. (2012). A time-varying effect model for intensive longitudinal data. *Psychological Methods*, 17(1), 61-77. doi: 10.1037/a0025814
- Tarvainen, M. P., Georgiadis, S. D., Ranta-aho, P. O., & Karjalainen, P. A. (2006). Time-varying analysis of heart rate variability signals with Kalman smoother algorithm. *Physiological Measurement*, 27, 225-239. doi: 10.1088/0967-3334/27/3/002
- Thatcher, R. W. (1998). A predator-prey model of human cerebral development. In K. M. Newell & P. C. M. Molenaar (Eds.), *Applications of nonlinear dynamics to developmental process modeling* (p. 87-128). Mahwah, NJ: Lawrence Erlbaum.
- Thom, R. (1993). *Structural stability and morphogenesis: An outline of a general theory of models*. Reading, MA: Addison-Wesley.
- Thomas, E. A., & Martin, J. A. (1976). Analyses of parent-infant interaction. *Psychological Review*, 83(2), 141-156. doi: 10.1037/0033-295X.83.2.141
- Trail, J. B., Collins, L. M., Rivera, D. E., Li, R., Piper, M. E., & Baker, T. B. (2013). Functional data analysis for dynamical system identification of behavioral processes. *Psychological Methods*, 19, 175-182. doi: 10.1037/a0034035
- Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods*. New York, NY: Academic Press.
- Uhlenbeck, G. E. (1980). Some notes on the relation between fluid mechanics and statistical physics. *Annual Review of Fluid Mechanics*, 12, 1-9. doi: 10.1146/annurev.fl.12

- .010180.000245
- Vallacher, R. R., & Nowak, A. (Eds.). (1994). *Dynamical systems in social psychology*. San Diego, CA: Academic Press.
- van der Maas, H. L. J., Kolstein, R., & van der Pligt, J. (2003). Sudden transitions in attitudes. *Sociological Methods & Research*, 32(125-152). doi: 10.1177/0049124103253773
- Voelkle, M. C., & Oud, J. H. L. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British Journal of Mathematical and Statistical Psychology*, 103-126. doi: 10.1111/j.2044-8317.2012.02043.x
- Volterra, V. (1926). Variazioni e fluttuazioni del numero di individui in specie animali conviventi (variations and fluctuations in the number of individuals in cohabiting animal species). *Mem. Acad. Lincei Roma*, 2, 31-113. doi: 10.1038/118558a0
- Wagenmakers, E.-J., Molenaar, P. C. M., Grasman, R. P. P. P., Hartelman, P. A. I., & van der Maas, H. L. J. (2005). Transformation invariant stochastic catastrophe theory. *Physica D*, 211, 263-276. doi: 10.1016/j.physd.2005.08.014
- Weiss, A. A. (1985). The stability of the AR(1) process with an AR(1) coefficient. *Journal of Time Series Analysis*, 6, 181-186. doi: 10.1111/j.1467-9892.1985.tb00408.x
- West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.). New York, NY: Springer-Verlag. doi: 10.1007/b98971
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In *Handbook of structural equation modeling* (p. 209-231). New York, NY, US: Guilford Press. doi: 10.4135/978-1-8570-2099-1.n23
- Witkiewitz, K., van der Maas, H. L. J., Hufford, M., & Marlatt, G. A. (2007). Nonnormality and divergence in post-treatment alcohol use: Reexamining the project MATCH data "another way". *Journal of Abnormal Psychology*, 116, 378-394. doi: 10.1037/0021-843x.116.2.378
- Wood, F. S. (1973). The use of individual effects and residuals in fitting equations to data. *Technometrics*, 15(4), 677-695. Retrieved from <http://www.jstor.org/stable/1267381> doi: 10.1080/00401706.1973.10489104
- Wood, S. (2018). Package 'mgcv' [Computer software manual]. Retrieved from <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf> (R package version 1.8-23)
- Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, 65(1), 95-114. doi: 10.1111/1467-9868.00374
- Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: CRC press. doi: 10.1201/9781420010404
- Yerkes, R., & Dodson, J. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482. doi: 10.1002/cne.920180503
- Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, 234(4), 65-83. doi: 10.1038/scientificamerican0476-65
- Zill, D. G. (1993). *A first course in differential equations* (5th ed.). Boston: PWS-KENT Publishing Company.