



The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data

Nicholas C. Jacobson¹ · Sy-Miin Chow¹ · Michelle G. Newman¹

© Psychonomic Society, Inc. 2018

Abstract

With the recent growth in intensive longitudinal designs and the corresponding demand for methods to analyze such data, there has never been a more pressing need for user-friendly analytic tools that can identify and estimate optimal time lags in intensive longitudinal data. The available standard exploratory methods to identify optimal time lags within univariate and multivariate multiple-subject time series are greatly underpowered at the group (i.e., population) level. We describe a hybrid exploratory–confirmatory tool, referred to herein as the *Differential Time-Varying Effect Model (DTVEM)*, which features a convenient user-accessible function to identify optimal time lags and estimate these lags within a state-space framework. Data from an empirical ecological momentary assessment study are then used to demonstrate the utility of the proposed tool in identifying the optimal time lag for studying the linkages between nervousness and heart rate in a group of undergraduate students. Using a simulation study, we illustrate the effectiveness of DTVEM in identifying optimal lag structures in multiple-subject time-series data with missingness, as well as its strengths and limitations as a hybrid exploratory–confirmatory approach, relative to other existing approaches.

Keywords Lag · Time-series analysis · Vector autoregressive modeling · Multivariate analysis · Splines

Since the 1990s, studies utilizing intensive longitudinal data (ILD; e.g., ecological momentary assessment, experience sampling, or physiological measurements) have grown exponentially in popularity. For instance, Google Ngram Viewer, an online phrase-usage graphing tool, suggests that publications with the term “experience sampling assessment” had grown by 385% in 2008, relative to 1990 (Michel et al., 2011). Among the most popular ILD designs in the behavioral sciences are variations of the multivariate, multiple-subject, replicated time-series design (Nesselroade & Ford, 1985), wherein intensive repeated measures of multiple variables within relatively short time lengths are collected from multiple subjects (prototypically, about four or five times daily over the course of a week; Bolger & Laurenceau, 2013; Nesselroade & Ford, 1985; Shiffman, Stone, & Hufford, 2008).

Coinciding with its increased prevalence, there has been great progress in analyzing ILD (Walls & Schafer, 2006). This includes linear and nonlinear time series models (e.g.,

variations of vector autoregressive moving average [VARMA] models), state-space models, dynamical systems models, multilevel modeling, and various other examples (Boker & Graham, 1998; Box, Jenkins, & Reinsel, 2013; Browne & Nesselroade, 2005; Durbin & Koopman, 2012; Goldstein, Healy, & Rasbash, 1994). Importantly, the advancement of new methods has increased the variety of research questions that can be answered by ILD.

In a discrete-time modeling framework, one model variation of interest is a vector autoregressive model of order p [VAR(p) model], in which multivariate measurements of a set of endogenous (dependent) variables from up to p measurement occasions ago (i.e., $t-1, t-2, \dots, t-p$) are used to explain these variables’ current values at time t . VAR models continue to serve as an important basis for many substantive studies and new methodological innovations (Bringmann et al., 2017; Chow, Hamagani, & Nesselroade, 2007; Epskamp et al., 2018; Gates, Molenaar, Hillary, Ram, & Rovine, 2010). Values from previous occasions are referred to as *lagged* responses, with the number of lags defined as the number of backward shifts in measurement occasion such that a lag-0 response refers to the current response at time t , a lag-1 response refers to the response at time $t-1$, a lag-2 response refers to the response at time $t-2$, and so on. Relatedly, p in a

✉ Nicholas C. Jacobson
njacobson88@gmail.com

¹ Pennsylvania State University, University Park, PA, USA

VAR(p) model is denoted as the *lag order* of the VAR process. The VAR(p) model is expressed as

$$\mathbf{y}_{i,t} = \mathbf{A}_1 \mathbf{y}_{i,t-1} + \mathbf{A}_2 \mathbf{y}_{i,t-2} + \dots + \mathbf{A}_p \mathbf{y}_{i,t-p} + \boldsymbol{\varepsilon}_{i,t} \quad (1)$$

where $\mathbf{y}_{i,t}$ represents an $M \times 1$ vector of observed variables measured on occasion t , $\mathbf{A}_p$ is an $M \times M$ matrix of regression coefficients containing the autoregression and cross-regression coefficients from lag p (i.e., from time $t - p$) on the m observed variables at time t ; and $\boldsymbol{\varepsilon}_{i,t}$ is an $M \times 1$ vector of residuals (process noises). A VAR(p) model can also be conceived as a set of difference equations or equivalently, continuous-time differential equation models in which the rate of change, change in the rate of change, and other higher-order changes are defined as unfolding over a time interval, Δt , of 1.0 (Hamilton, 1994).¹ In practice, discrete-time models such as VAR(p) models have typically been fitted to data measured at discrete, mostly equally spaced intervals (aside from occasional missingness). Most often, the measurement interval in a study is assumed by design to correspond to $\Delta t = 1$ in the underlying difference or differential equation model.

The measurement intervals utilized for data collection have direct implications on the strengths of the regression coefficients linking the lagged responses to the current responses at time t —a point that has been brought up by several researchers in advocating direct use of continuous-time models over discrete-time models (Kuiper & Ryan, 2018; Voelkle & Oud, 2013). We will return to this point in the Discussion section. For now, suffice it to say that when researchers have equally spaced data, fitting discrete—as opposed to continuous—time to the data often has some practical advantages. For instance, a wide and well-established array of tools are available in the statistical and econometric literature for diagnosing, exploring, and interpreting results from discrete-time models, particularly those involving complex lag structures (e.g., when p is high, and/or when different processes show distinct optimal lag orders). Inference and interpretations of continuous-time models have thus far been limited to those involving a lower lag order, which, in a continuous-time framework, mirrors the highest-order changes included in a differential equation of choice (e.g., rate of change is a first-order change, acceleration/deceleration is a second-order change). Tools for model exploration are still nascent, and estimation of continuous-time models is often characterized by greater numerical difficulties than estimation of their discrete-time counterparts of the same order.

When a VAR(p) model is determined to be the model of choice and researchers have a priori knowledge concerning the true lag order, p , of the process, researchers can proceed directly to model fitting. However,

when the optimal lag order is unknown, the lag order is often determined using exploratory tools such as autocorrelation, partial autocorrelation, cross-correlation, and partial cross-correlation plots (Chatfield, 2013; Turchin & Taylor, 1992).² Because these diagnostic methods originated primarily from the time series or econometric literature, current software implementation of the autocorrelation, partial autocorrelation, cross-correlation, and partial cross-correlation functions is restricted for use with single-subject time series, univariate or bivariate data, and often assuming no missingness. When multiple endogenous variables are present, selecting the optimal lag order for each process and that for all the processes as a system can quickly become a cumbersome variable selection problem: not all processes would have the same p , and not all the coefficients linked to the $t - 1, t - 2, \dots, t - p$ lagged responses of the same variable (termed *autoregression coefficients*) and of different variables have to be freed up for reasons of parsimony.

To illustrate the utility of the partial autocorrelation and partial cross-correlation functions in the scenario for which they were originally intended (Turchin & Taylor, 1992), we first present diagnostic plots of the partial autocorrelation with an autoregressive model of order 1 [a VAR(1) model; termed *Illustration 1*, which is based on Simulation 1 below], one person, 100 time points, and complete data. In particular, the partial autocorrelation function, which summarizes the correlations of a variable with itself at various lags after the effects of lower-order lags have been partialled out, correctly identifies the autoregressive lag of 1 with no Type I errors (Fig. 1). In our second illustration (i.e., *Illustration 2*), we present diagnostic results from brief simulation based on a bivariate model of order 14 [a VAR(14) model], with simulated nonzero autoregression coefficients at lags 1, 3, 4, 12, and 14 for one variable; nonzero autoregression coefficients at lags 1 and 3 for the second variable; and nonzero cross-regression coefficients at lags 8 and 12 for the first variable on the second variable (see the empirical example below for the motivation, as well as the multivariate case in our Simulation 3 below). With simulated data from a single subject over 170 time points and no missingness, the diagnostic results based on the partial autocorrelation and partial cross-correlation functions are shown in Fig. 2. In particular, the partial autocorrelation function estimated three of the seven nonzero autoregressive lags as significant. However, it lacked power to detect the remaining lags, whose coefficients were characterized by smaller (absolute) magnitudes. The partial cross-correlation function (see Wei, 2006, pp. 402–414), which serves to detect correlations of a variable with another variable at prespecified

¹ Using the analogy of distance traveled by a car, the rate of change corresponds to the instantaneous speed of the car, the change in the rate of change refers to the current acceleration or deceleration of the car, and so on.

² For this article, the term “optimal lag” is defined as the lags that best predict anticipated outcomes. This term is used when the lag structure in the population is unknown (i.e., in normal empirical situations). The term “true lag order” is the lag order in the population.

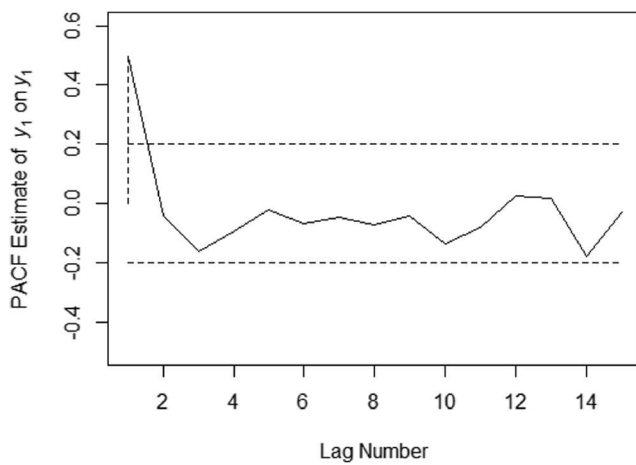


Fig. 1 This plot depicts the partial autocorrelation function results with single-subject data, 100 time points, and complete data. The solid black line represents the partial autocorrelation function estimates; the dashed black lines represent the critical values [based on $\pm 2/\sqrt{(T-1)}$, where T = the number of time points, 100 in this case]. In this plot, the lag of 1 is correctly identified as being significantly different from 0. In addition, there are no Type I errors (a.k.a. false positives). This is the scenario for which the partial autocorrelation function was designed.

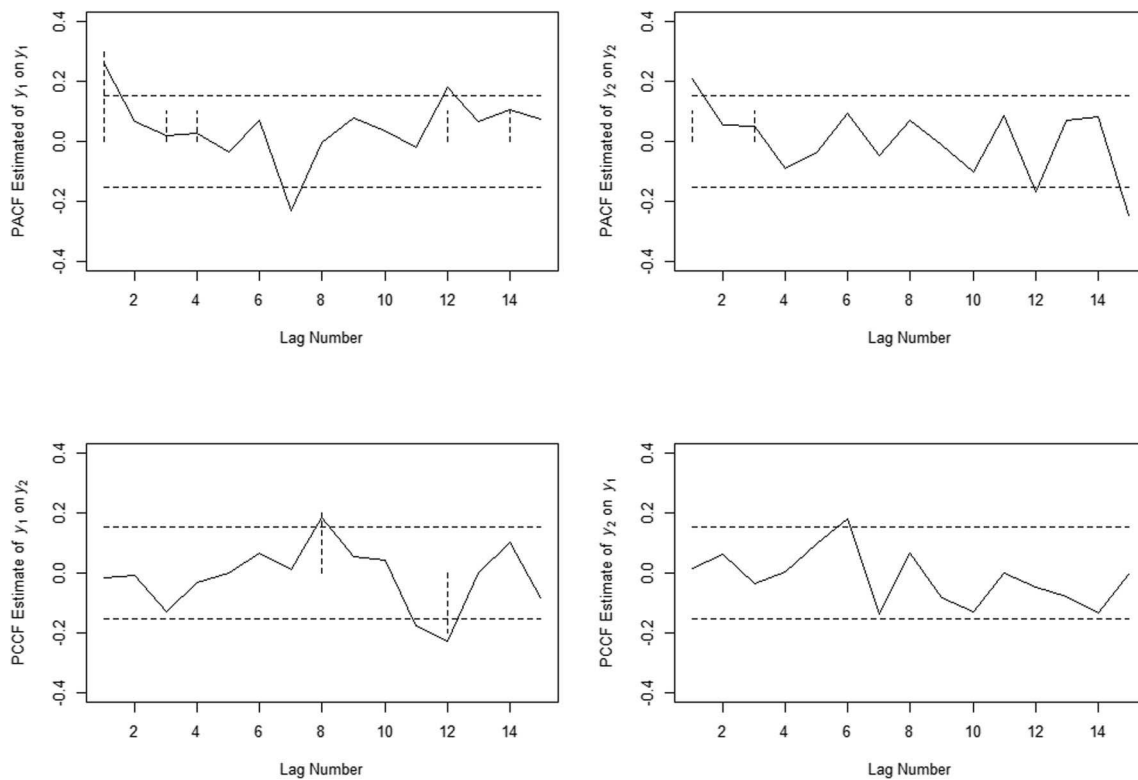


Fig. 2 These plots depict the partial autocorrelation and partial cross-correlation function results with single-subject data, 170 time points, and complete data. The solid black lines represent the partial autocorrelation function estimates; the dashed black lines represent the critical values [based on $\pm 2/\sqrt{(T-1)}$, where T = the number of time points, 170 in this case]. This figure depicts from left to right, top to bottom: (top left) the partial autocorrelation function results of y_1 on itself (lags 1 and 12 are correctly identified as significant, but lags 3, 4, and 14 are not identified due to their low magnitude, and lag 7 is incorrectly identified

lags after the effects of lower-order lags have been partialled out, correctly identified both nonzero cross-regressive lags (although also two lags were incorrectly identified as significant). These results suggest that although these methods might be underpowered in low-effect-size scenarios, these tools performed adequately in identifying lags with higher strength in complete single-subject time series (i.e., their intended function).

Next, we focused on scenarios that were explicitly beyond the intended purpose of these exploratory functions, with the sort of data typically seen in the behavioral sciences (Turchin & Taylor, 1992). Particularly, we focused on the performance of these methods in estimating group-level statistics from multiple subjects with missing data. In a third simulation based on the same simulation model in Illustration 2, we tested the utility of these exploratory functions for situations that are typical of daily-diary data in the behavioral sciences (Roche, Jacobson, & Pincus, 2016), in which there were a large number of subjects and short time series. We retained complete data in this simulation to show that even without missing data (which is quite uncommon

as significant); (top right) the partial autocorrelation function results of y_2 on itself (lag 1 is correctly identified as significant, but lag 3 is not identified as significant, and lag 15 is incorrectly identified as significant); (bottom left) the partial cross-correlation function results of y_1 on y_2 (lags 8 and 12 are correctly identified as significant, and lag 11 is incorrectly identified as significant); and (bottom right) the partial cross-correlation function results of y_2 on y_1 (where lag 6 is incorrectly identified as significant). This is the scenario for which the partial autocorrelation function and partial cross-correlation function were designed.

in ILD within the behavioral sciences), the partial autocorrelation and partial cross-correlation functions were not adequate for such designs. With 159 subjects, 14 time points, and complete data (i.e., *Illustration 3*), the results were substantially worse than for the single-subject simulation (see Fig. 3). In particular, no lags were identified as significant, and there were no clearly recognizable patterns even when collapsing across the 159 different partial autocorrelation and partial cross-correlation estimates. This showed that the partial autocorrelation and partial cross-correlation functions did not perform adequately in multiple-subject data with short time series.

Fourth, we demonstrated that the partial autocorrelation and partial cross-correlation functions performed poorly with single-time-series data when there were missing data (i.e., *Illustration 4*). In simulating single-subject data with 72% of the data missing (which was the level of missing data in the empirical example),

the model incorrectly identified 24 lags as significant (i.e., 24 Type I errors) and failed to identify seven of nine nonzero lags as significant in the correct direction (see Fig. 4). Further details of these demonstrations are provided later, in the context of a simulation study (see Simulation 3). For now, they serve to highlight the inadequacies of current exploratory methods in correctly identifying potential lags in multiple-subject time-series data of finite lengths and with missing data—namely, the kind of ILD that are typically available in psychology and other behavioral sciences.

One alternative strategy is to simply ignore higher-order lags due to convenience or due to theory, and simply to focus on lower-order lags (most typically, just lags of 1). This strategy assumes that the effects of higher-order lags are negligible as compared to those of lower-order lags. In other words, there are stronger associations among measurement occasions that are more closely separated in time than among those that are

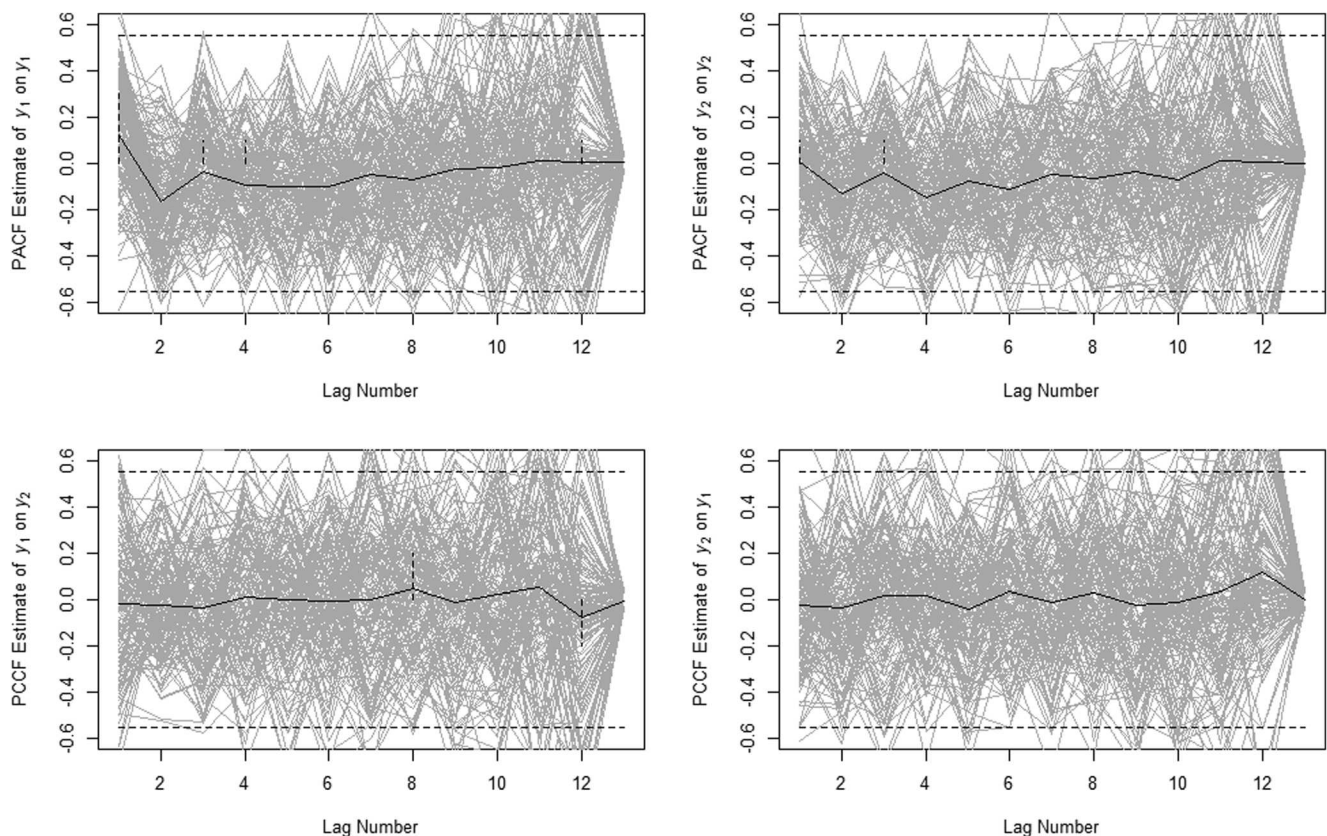


Fig. 3 These plots depict the partial cross-correlation function results with multiple-subject data, 14 time points, and no missing data. The gray lines in this figure depict the results of each of the individual cross-correlation functions. The solid black lines represent the averages of the individual cross-correlation function estimates, to obtain a group-level autocorrelation function estimate. The dashed black lines represent the critical values [based on $\pm 2/\sqrt{(T-1)}$, where T = the number of time points, 14]. This figure depicts, from left to right, top to bottom: (top left) the partial autocorrelation function results of y_1 on itself; (top right) the partial autocorrelation function results of y_2 on itself; (bottom left) the

partial cross-correlation function results of y_1 on y_2 ; and (bottom right) the partial cross-correlation function results of y_2 on y_1 . In all plots, no lags are correctly identified as significant. This demonstrates that the partial autocorrelation and partial cross-correlation function performs quite poorly with multigroup data, because the formula for the 95% confidence lines utilizes T from a single subject and does not pool information across multiple subjects. Note that if N is added to the critical value formula $\pm 2/\sqrt{(T*N-1)}$, most of the simulated nonzero lags would still not be significant, and many spurious lags would be identified.

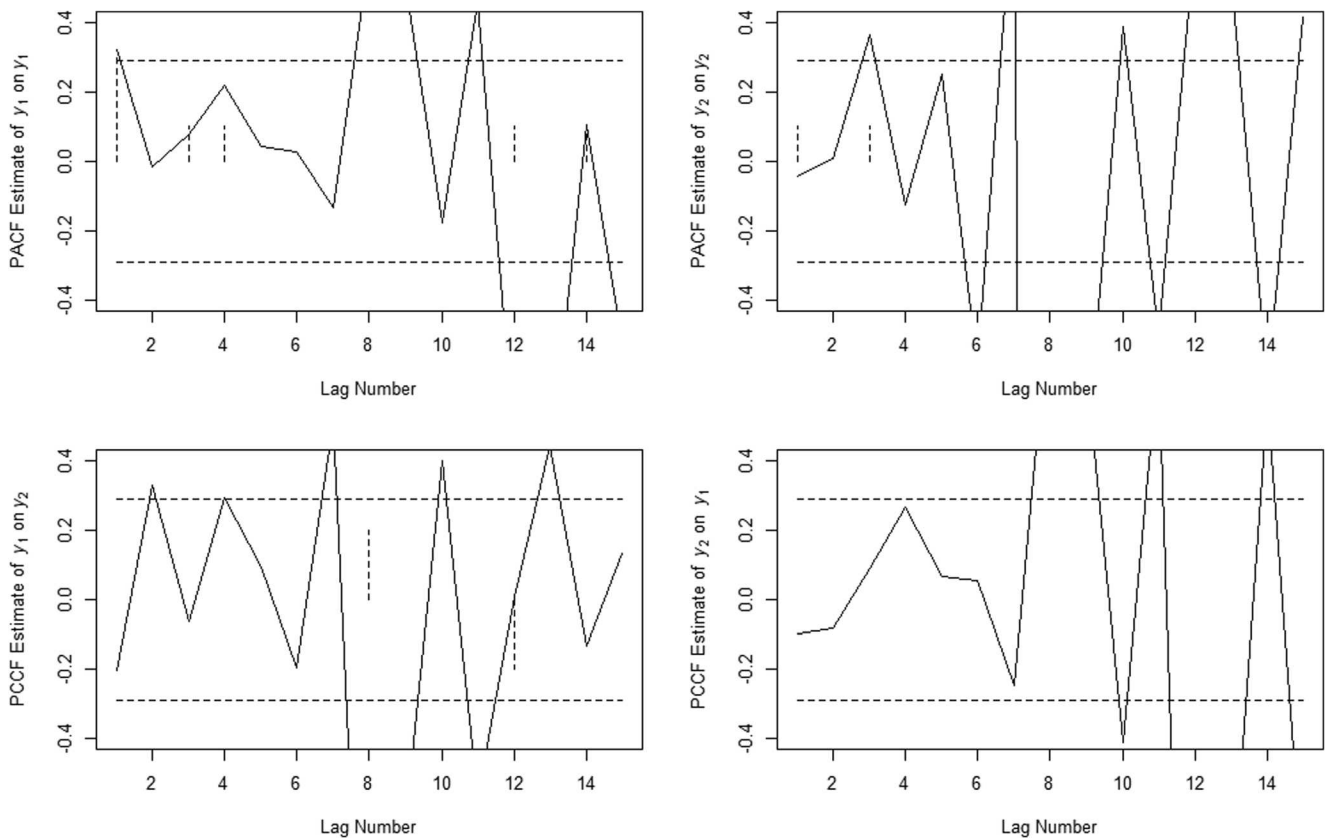


Fig. 4 These plots depict the partial cross-correlation function results with single-subject data, 170 time points, and 72% of the data missing. The solid black lines represent the partial cross-correlation function estimates. The dashed black lines represent the critical values [based on $\pm 2/\sqrt{(T-1)}$, where T = the number of time points, 48]. This figure depicts, from left to right, top to bottom: (top left) the partial autocorrelation function results of y_1 on itself; (top right) the partial autocorrelation function results of y_2 on itself; (bottom left) the

partial cross-correlation function results of y_1 on y_2 ; and (bottom right) the partial cross-correlation function results of y_2 on y_1 . In these plots, only lag 1 for y_1 on y_1 and lag 3 for y_2 on y_2 are correctly identified as significant. Of the other seven simulated nonzero lags, two were identified as significant, but in the incorrect direction, and 24 Type I errors are made. This demonstrates that the partial autocorrelation and partial cross-correlation functions both perform quite poorly at the single-subject level when there are missing data.

farther apart—not an unreasonable assumption for many processes in the social and behavioral sciences. However, as we next illustrate, the impact of failing to incorporate higher-order lags when they exist may not be as innocuous as is commonly assumed. Consider the same VAR(14) model in our first illustration, but now with 72% missing data (see Simulation 3). If one simply ignores all higher-order lags beyond lag 1 and proceeds with estimating a VAR(1) model, inaccurate point estimates would result. In particular, when ignoring higher-order lags, the 95% confidence intervals for the autoregression coefficient of lag 1 for y_1 did not include the simulated population value. Moreover, the cross-regressive lag of y_2 on y_1 at lag 1 was identified as significant. In contrast, when the data were fitted with a correctly specified lag structure, the point estimates were all extremely close to their simulated values (see the results of Simulation Study 3 below for greater detail). Thus, in addition to obtaining an incomplete test of one's theories, ignoring higher-order lags can lead to misleading inferential results.

The present article describes a convenient, user-accessible exploratory and confirmatory tool to find the optimal lag structure in VAR(p) models (Browne & Nesselroade, 2005; Nesselroade, McArdle, Aggen, & Meyers, 2001). In particular, the present procedure allows users to first narrow the search space estimated in VAR(p) models by using variable selection methods grounded within the penalized additive modeling framework (Wood, 2003, 2006). Within the penalized additive modeling framework, each of the potential lags in the search space is translated into a large number of independent variables or manifest predictors (e.g., Illustration 3, with 2 predictors $\times$ 2 dependent variables $\times$ 14 lags considered would result in a total of 56 manifest independent variables). Nevertheless, variable selection methods in the penalized additive framework allow one to reduce the search space by down-weighting negligible (near-zero) lag coefficients to zero (similar to Gasparrini, Armstrong, & Kenward, 2010). After determining the optimal lag structure [including information on p in a VAR(p) model, in conjunction with the auto- and cross-regression coefficients that

are to be freely estimated vs. constrained to be zero], the results from fitting the final selected VAR(p) model in a confirmatory framework are returned. Ultimately, the primary utility of this approach is to offer users a convenient way of exploring and modeling optimal lag structures in a VAR(p) framework.

In addition to describing the novel combination of these two methods for modeling frameworks in identifying and modeling the optimal lag structure, the present article provides an empirical illustration and simulation studies demonstrating the utility, strengths, and limitations of these techniques.

Differential Time-Varying Effect Model (DTVEM)

DTVEM is a set of integrated subroutines for diagnosing the optimal lag identification in equally spaced ILD for single or multiple subjects.³ In particular, DTVEM combined some of the flexible smoothing and estimation routines for fitting generalized additive mixed models (GAMMs, available as part of the R package mgcv; Wood, 2006) and the state-space estimation routines available in the R package OpenMx (Chow, Ho, Hamaker, & Dolan, 2010; Harvey, 2001; Neale et al., 2016) to explore, diagnose, and fit group-based VAR models of unknown lag structures. That is, we assumed that the underlying model of interest was a group-based VAR model with the same lag structure across all individuals, but this optimal lag structure was unknown and had to be detected using multivariate, replicated time series from multiple subjects.

The exploratory stage using GAMM was useful in narrowing down the search space, as lags that had substantial effects on outcome variable(s) were weighted heavily, and lags that had negligible effects on outcome variable(s) were down-weighted. In “smoothing over” the effects of successive lags, this approach provides a parsimonious but flexible way to explore relationships among multiple variables while considering possible lagged associations among them. Using GAMM to narrow down the search space prior to confirmatory model estimation can be important, as large numbers of lags can be computationally inefficient, and interdependencies from concurrent and lagged associations can sometimes interfere with confirmatory estimation (e.g. Botta & Perez Truglia, 2011; Buckner, Crosby, Wonderlich, & Schmidt, 2012; Buysse et al., 2007; Carels et al., 2004; Starr & Davila, 2012). Following this exploratory stage with a confirmatory stage within the state-space framework subsequently allows one to estimate the identified lags with greater precision than at the exploratory stage. The integrated routines cycle iteratively among exploratory lag detection and confirmatory model fitting and output maximum-likelihood estimates

from the “final” model in the final iteration. These automated routines were labeled *using the DTVEM function* shown in the Appendix. Because they utilized subroutines developed for both the GAMM and state-space frameworks, we describe each of these in turn.

Narrowing down the search space with the generalized additive mixed effects model (GAMM)

To narrow down the search space for the confirmatory stage, DTVEM offers users the choice to first explore the lag structure in the GAMM framework prior to proceeding to the confirmatory stage. The general GAMM framework, which extends the generalized linear model (McCullagh & Nelder, 1989) and generalized additive model (Hastie & Tibshirani, 1993; James, 2002), postulates that person i 's any one particular response variable, y_i (where $i = 1, \dots, n$; n indexes the total number of subjects), may be distributed as any of the members from the exponential family (e.g., normal, Poisson, gamma, multinomial, etc.; for further examples, see chap. 13 of Cohen, Cohen, West, & Aiken, 2003). The mean of y_i , $\mu_i \equiv E(y_i)$, is linked to a semiparametric predictor, η_i , expressed as

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_{1,k}(x_{1,k,i}) + \sum_{o=1}^O \sum_{c=1}^C f_{2,c,o}(x_{2,o,i}) x_{2',c,i} + \sum_{q=1}^Q \sum_{s=1}^S f_{3,s,q}(x_{3,s,i}, x_{3',s,i}) + \mathbf{Z}_i \mathbf{b}_i \quad (2)$$

via $\eta_i = g(\mu_i)$, where g is a link function that maps the mean of y_i to η_i , and $g^{-1}(\eta_i)$ is the reverse transformation that converts η_i into μ_i . The first and last terms constitute the usual parametric components in standard linear mixed effects models; and the second, third, and fourth terms are nonparametric components (i.e., of unknown functional forms) wherein the effects of a series of covariates on the mean of the transformed dependent response variable are of unknown functional forms. Specifically, $\mathbf{X}_i$ is a $1 \times n_\beta$ design vector that contains person i 's fixed-effects components; $\boldsymbol{\beta}$ is the corresponding $n_\beta \times 1$ vector of fixed-effects parameters; $\mathbf{Z}_i$ is the $1 \times n_b$ random-effects design vector for person i ; and $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\psi}_b)$ is a vector of random effects, assumed to be multivariate normally distributed with zero means and positive-definite covariance matrix $\boldsymbol{\psi}_b$. The terms $f_{1,k}(\cdot)$, $f_{2,c,o}(\cdot)$, and $f_{3,s,q}(\cdot)$ are nonparametric functions involving the covariates in different ways.⁴ The covariates that appear

³ These integrated routines have been compiled into a working R package that can be downloaded via the following Web address: www.nicholasjacobson.com/projects/dtvem.

⁴ GAMM provides a collection of procedures for approximating these functions and the resultant curves using different smoothers. $f_{1,k}(\cdot)$ – $f_{3,s,q}(\cdot)$ are typically referred to as *smooth functions*, and the curves or lines produced by these functions are denoted as *smooths* (Hastie & Tibshirani, 1990; McKeown & Sneddon, 2014). Note that because the number of basis functions used with each covariate can generally be different, the nonparametric part cannot be written in matrix form because the dimensions of the basis coefficient vectors and basis function vectors generally differ between covariates.

in each smooth function may be different, and we thus use different indices and subscripts to distinguish the covariates that appear in the three sets of smooth functions.

The term $f_{l,k}(\cdot)$ is the smooth function of the k th covariate, $x_{1,k,i}$ ($k = 1, \dots, K$). For instance, if y_i represents depression for person i and $x_{1,k,i}$ represents the anxiety of person i , $f_{l,k}(\cdot)$ then captures the unknown association between depression and anxiety across all the individuals in the sample. This is specified as $s(x_{1,k,i})$ in the model specification portion of the *gamm* function. Because a nonparametric function is used, the resultant smooth (approximation curve) may be linear or nonlinear. The third term allows the smooth function for the o th covariate, $f_{2,c,o}(x_{2,o,i}, c = 1, \dots, C; o = 1, \dots, O)$, to depend on, or interact with another (unsmoothed) covariate $x_{2',c,i}$ (Ahmad, Leelahanon, & Li, 2005; Hastie & Tibshirani, 1993). In GAMM, the unsmoothed covariate is specified in the model specification line using the “by” keyword to specify a varying-coefficient model (Ahmad et al., 2005; Chow, Zu, Shifren, & Zhang, 2011; Hastie & Tibshirani, 1993; Shiyko, Lanza, Tan, Li, & Shiffman, 2012), in which the effect of the c th covariate is assumed to vary nonparametrically but smoothly over another covariate such as time or geographical regions. Finally, the term $f_{3,s,q}$ ($s = 1, \dots, S; q = 1, \dots, Q$) is a tensor product used to approximate the unknown but possibly jointly nonlinear effects of a pair of covariates on η_i .

Typically, when researchers fit VAR-type models in a regression framework, one approach is to create p new lagged variables for the intended lag order. So for four lags, four new variables will have to be included in the data set and model (e.g., Chow, Haltigan, & Messinger, 2010). However, this requires a priori decision on p , the maximum order in the VAR(p) process. Specifying an arbitrarily large value for p , in contrast, would lead to a large number of independent variables and correspondingly, missingness in these independent variables. Using GAMM for this exploratory stage, we estimate one lag variable that contains all possible lags; differences across lags are only distinguished by means of a covariate, time differences (i.e. Δt). Thus, at this exploratory stage, we fit lags using the varying-coefficient model in GAMM [using the third term, $f_{2,c,o}(x_{2,o,i}, c = 1, \dots, C; o = 1, \dots, O)$, and thereby estimate the linear association between lagged variables and the outcome as a smooth function of nonlinear time differences.

See Fig. 5 for an illustration of this exploratory stage when used to detect significant lagged associations using simulated data from a model with nonzero lags of 8 and 12 (see the multivariate example for greater details). These lags are evident as peaks and valleys whose 95% confidence intervals did not include zero at time differences of 8 and 12.⁵ These plots

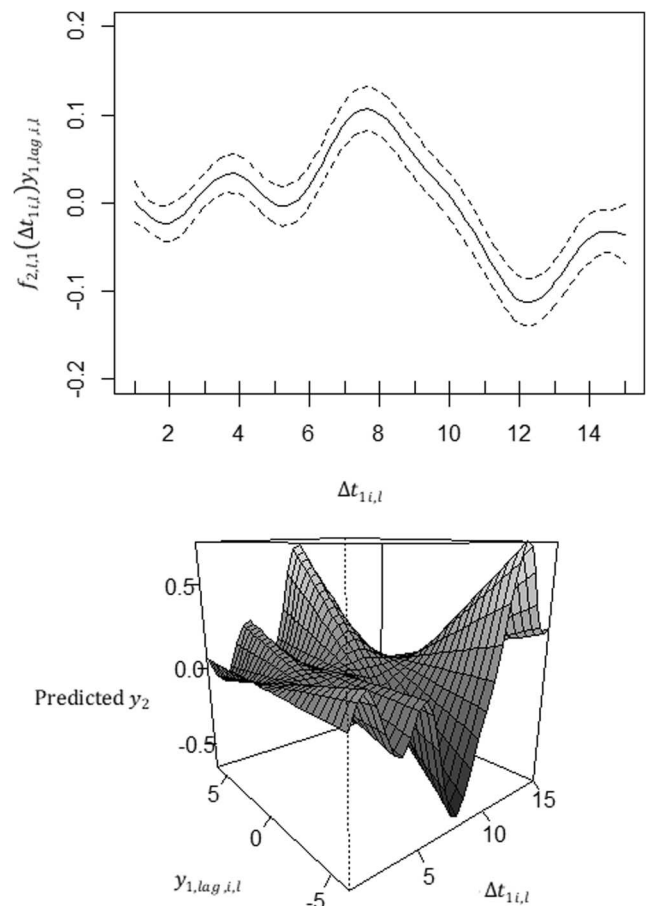


Fig. 5 These figures present the output from the exploratory stage of DTVEM. In the first figure, the solid black line represents the predicted values, and the dashed lines represent the corresponding 95% confidence intervals. The data-generating model was simulated to have nonzero lags of 8 and 12. The label $f_{2,l,1}(\Delta t_{1,l})y_{1,lag,i,l}$ depicts the smoothed weight of $y_{1,lag,i,l}$ on $y_{i,l}$ at every possible value of $\Delta t_{1,l}$.

also highlight one property or limitation of using the exploratory stage of DTVEM alone. That is, in VAR processes—the kind of group-based process modeled in DTVEM at the moment—the effects of earlier lags would still linger at later lags unless the effects of the earlier lags have been partialled out. This “limitation” is similar in nature to the inadequacy of diagnosing lag structures in AR processes using auto- as opposed to partial autocorrelation functions. Thus, results from the exploratory step cannot be interpreted alone, but rather lags identified in the exploratory stage are passed on to the confirmatory state-space framework.

Social and behavioral scientists are often interested in the simultaneous estimation of outcomes within a single model to investigate lead-lag relationships among multiple dependent variables. In this case, Eq. 2 can be extended to involve M dependent variables, using the dummy variable indicator approach by MacCallum, Kim, Malarkey, and Kiecolt-Glaser (1997).

Because the linear state-space model is parametric in nature, nonparametric (possible nonlinear) time trends identified

⁵ Note that local peaks are determined by examining the neighboring values and determining whether the coefficient of interest is greater than these values, and if the coefficient is also significantly greater than 0. Each local valley is the inverse of a local peak.

in this exploratory stage are first removed (yielding group-detrended data) before proceeding to the confirmatory model.

In sum, estimating the associations between the predictor and the outcome across a smooth of non-linear time differences help one to detect optimal lag structures. Moreover, this smoothing procedure helps to retain only the important lags for the confirmatory stage.

Confirmatory stage: Verification of lag structure via a confirmatory VAR(p) model

The primary utility of DTVEM is estimating confirmatory VAR(p) models (see Eq. 1). As an example, consider a VAR(3) with two dependent variables, which would be expressed as

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \mathbf{A}_1 \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + \mathbf{A}_2 \begin{bmatrix} y_{1,i,t-2} \\ y_{2,i,t-2} \end{bmatrix} + \mathbf{A}_3 \begin{bmatrix} y_{1,i,t-3} \\ y_{2,i,t-3} \end{bmatrix} + \boldsymbol{\varepsilon}_{i,t} \quad (3)$$

where $y_{1,i,t}$ and $y_{2,i,t}$ represent the vector of observed dependent variables measured on occasion t . The matrices $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_3$ are each 2×2 matrices in this particular example, with the diagonal entries corresponding to autoregression coefficients for the two dependent variables at lags 1, 2, and 3, respectively, whereas the off-diagonal entries represent the cross-regression coefficients at those lags.

If the user opts to use the exploratory stage, not all of these auto- and cross-regression coefficients illustrated in Eq. 3 would need to be freed up and estimated all at once. Rather, subsets of these coefficients, as identified to be significant within the exploratory stage, are freed up and estimated. Alternatively, the user can opt to estimate all auto- and cross-regressions if they so desire. Our particular confirmatory approach of specifying a VAR(p) process with known lag structure as a state-space model and obtaining the associated maximum-likelihood parameter estimates by optimizing the so-called “prediction error decomposition function” is known to yield satisfactory point and standard error estimates when the correctly specified model is fitted (Chow, Ho, et al., 2010; Harvey, 2001).

Additional exploratory and confirmatory iterations

Following the confirmatory stage, the user is given the choice to repeat additional exploratory stages to ensure that all potentially statistically significant lags are identified while controlling for the effects that were significant in the confirmatory stage. If the user opts for additional exploratory stages, the varying-coefficients from the first exploratory iteration are nearly identical. However, lagged responses at particular lags found to yield statistically significant coefficients at the confirmatory stage were included as additional columns in X_i in Eq. 2 and removed from

the list of predictors with varying coefficients, $f_{2,c,o}(\cdot)$ in Eq. 2. This additional exploratory stage essentially partials out the effects from those particular lags before other lags are evaluated again. Any newly identified lags are then iteratively re-estimated with the confirmatory stage. If this is elected by the user, this process is repeated iteratively until no new peaks or valleys are identified. At the end of the DTVEM function, DTVEM outputs the last confirmatory state-space estimates.

All the stages of the DTVEM models are fully automated in the R function illustrated in the Appendix. In practice, a variety of spline functions or penalized spline functions may be used to obtain the smooths [i.e., all terms involving $f(\cdot)$] in these equations. In DTVEM, we use the thin-plate regression splines, which are a generalization of natural cubic splines (Bookstein, 1989), meaning that smooth curves are constructed from a series of higher-order polynomials with specific constraints that these functions must be smooth (Wood, 2003). Thin-plate regression splines use an eigenvalue decomposition to pick the basis coefficients that can explain the greatest variance. Thin-plate regression splines have the advantages of (1) not requiring a researcher to choose knot locations, thereby reducing subjectivity in modeling and otherwise having optimal bases (Wood, 2006) and (2) better accommodating a high number of predictors than other spline regression methods.

In summary, DTVEM was built to allow users to easily explore lags while reducing the search space by utilizing varying-coefficient models in GAMM and next using confirmatory state-space models to estimate optimal lag structures in the VAR framework.

Empirical example

The following example demonstrates the utility of DTVEM using an empirical problem involving the time course of anxious symptoms in daily life. Anxious moods are often thought to constitute the co-occurrence of several symptoms, including both feelings and physiological activation (Lang, McTeague, & Bradley, 2016). Recent movements have been particularly focused on the examination of these symptoms within one integrated framework (Cuthbert & Insel, 2013).

In contrast to theories about anxiety and physiological activation being concomitant with one another, there is considerable evidence that physiological measurements and self-reported anxious moods often have low associations when measured concurrently (Hodges, 2015; Mauss, Wilhelm, & Gross, 2004; Morris & Liebert, 1970). Rather than representing a single unitary construct, recent theories suggest that anxiety may be apprehended through multiple subsystems

(e.g., somatic arousal, anticipation of physiological arousal, and the avoidance of arousal are all considered integrated subsystems) that are connected to one another over different timescales (Epskamp et al., 2018; Frank, Jacobson, Hurley, & McKay, 2017). Although quite recent, such theories of a temporal relationship between physiology and subjective anxiety build on classical conceptualizations that physiological responses and the perception of emotion precede and predict one another over hours in a day (Cannon, 1927; Lange & James, 1922). Importantly, the interaction between physiology and feelings of anxiety are thought to inform the manifestation and maintenance of anxiety disorders (Frank et al., 2017). Consequentially, this relationship is crucial to the study of the nature of psychopathology within daily life.

In line with such theories, there is evidence that cognitive anxiety processes and negative emotions are associated with physiological activation later in the day and during the subsequent night (Brosschot & Thayer, 2003; Brosschot, Van Dijk, & Thayer, 2007). There also has been a suggestion that this may be due to the effects of anxiety on the neuroendocrine system, which can lead to prolonged heart rate over the span of hours (Mergler & Valciukas, 1998). Nevertheless, the optimal time in which the perception of anxiety and physiological activation predict one another remains unknown (Epskamp et al., 2018).

To date, no researcher has examined lead-lag relationships between perceptions of anxious moods and physiological reactions over the course of hours (Barrett, Quigley, Bliss-Moreau, & Aronson, 2004). Studying the temporal course of anxious moods over hours or even days in conjunction with lagged changes in physiological responses, such as heart rate, could hasten understanding of the phenomenology of anxious moods and have implications regarding the intersection of multiple units of analysis of pathological systems (Cuthbert & Insel, 2013). For example, if self-reported anxiety predicts later heart rate, treatment might focus on cognitive or emotional-processing therapy to prevent later increases in heart rate (Frank, Jacobson, Hurley, & McKay, 2017). In contrast, if physiological activation is a “leading indicator” of the perception of self-reported anxiety, it may suggest that anxious moods may be due to noticing changes in physiological responses (Epskamp et al., 2018). In this case, treatment might focus on relaxation techniques to directly target the physiology.

The present empirical example was based on a set of ecological momentary assessment data collected every hour that subjects were awake. Most participants ($N = 159$) completed 68.5% of prompts. The Profile of Mood States (POMS) “nervous” item was used on a 0–100 slider. Heart rate was measured with an open-source application that used the camera on smartphones

(Wetherell, 2013). The application used photoplethysmographic signals that were obtained by taking pictures of the color changes in the index finger when the finger was pressed against the phone’s camera. The application ran for 30 s, and average heart rate was measured during this time. This method of obtaining heart rate through smartphone applications has been validated (Scully et al., 2012) and had high convergence with traditional measures ($r = .98$ – 1.00 with heart rate; Bolkhovskiy, Scully, & Chon, 2012). For this illustrative data, the lack of a priori knowledge about the time lagged associations in the data motivated our use of DTVEM.

Since participants were prompted once per hour, the data were broken down into hourly segments for the analyses. A total of 7,509 data points were collected, out of a possible 26,880 (if prompts had occurred evenly and each person had a complete sampling of every period). Thus, on the basis of the sampling period of interest, the data were 28% complete. Because there was a theorized bidirectional relationship between nervousness and heart rate in the literature, a bivariate model was considered.

Note that the exploratory stage of DTVEM was fitted using the following equation:

$$\begin{aligned}\eta_{i,l}^* = & f_{1,1}(t_{i,l}^*) + f_{2,l,1}(\Delta t_{1i,l})NRonNR_{i,l}^* \\ & + f_{3,l,1}(\Delta t_{1i,l})NRonHR_{i,l}^* + f_{3,l,1}(\Delta t_{1i,l})HRonNR_{i,l}^* \\ & + f_{4,l,1}(\Delta t_{1i,l})HRonHR_{i,l}^*\end{aligned}$$

$NRonNR$ represents the autoregressive effect of nervousness, $NRonHR$ represents the cross-regressive effect of nervousness on heart rate, $HRonNR$ represents the cross-regressive effects of heart rate on nervousness, and $HRonHR$ represents the autoregressive effects of heart rate. Among the results available from the exploratory stage were nonparametric time trends at the group level, shown in Fig. 6. In regard to the time trends, nervousness showed significant time trends, that were highest at the first hours in the study and peaked

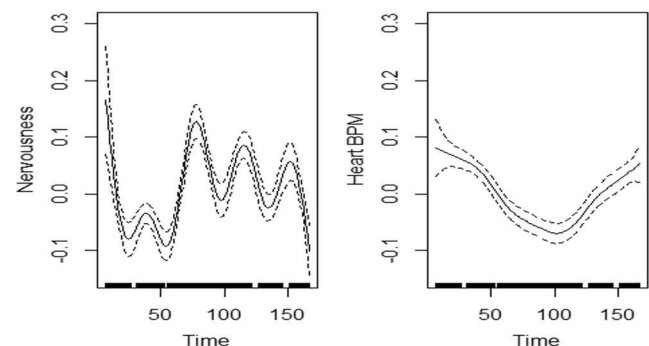


Fig. 6 These figures depict the time trends of nervousness and heart rate over time for the population.

again approximately 60 h later. Interestingly, heart rate began to decrease at this time. Notably, participants arrived for a compliance check at approximately this time, and this systemic increase in nervousness may coincide with the evaluation of their compliance.

The exploratory stage showed significant AR effects of nervousness and heart rate, such that the strongest effects of nervousness on nervousness occurred one to four hours later and peaked again 14 h later. The exploratory stage of DTVEM showed that the strongest AR effects of heart rate on itself occurred 1–3 h later. With regard to cross-regressive trends, the exploratory stage of DTVEM suggested that nervousness may positively predict heart rate 7 and 8 h later and negatively predict heart rate 12 h later. The exploratory stage of DTVEM suggested that heart rate might positively predict nervousness 15, 16, and 17 h later. See Fig. 7 for a summary of the exploratory-stage varying coefficients.

Following the optional exploratory stage, time trends were modeled by focusing solely on nervousness and heart rate time smooths. The residuals of this model were then passed on to the state-space confirmatory stage of DTVEM.

The final state-space confirmatory results indicated that nervousness predicted itself 1 h later ($\alpha_{1,1} = 0.275$, $SE_{\alpha_{1,1}} =$

0.012 , $p_{\alpha_{1,1}} < .001$), 3 h later ($\alpha_{1,3} = 0.071$, $SE_{\alpha_{1,3}} = 0.014$, $p_{\alpha_{1,3}} < .001$), 4 h later ($\alpha_{1,4} = 0.075$, $SE_{\alpha_{1,4}} = 0.014$, $p_{\alpha_{1,4}} < .001$), 12 h later ($\alpha_{1,12} = 0.077$, $SE_{\alpha_{1,12}} = 0.020$, $p_{\alpha_{1,12}} < .001$), and 14 h later ($\alpha_{1,14} = 0.082$, $SE_{\alpha_{1,14}} = 0.019$, $p_{\alpha_{1,14}} < .001$). The many significant AR terms associated with nervousness may suggest that the dynamics of nervousness are complex and conform the multiple time scales. These significant AR terms may also reflect person-specific time scales and individual differences, which were not modeled. In contrast, heart rate significantly predicted itself 1 h ($\alpha_{2,1} = 0.144$, $SE_{\alpha_{2,1}} = 0.015$, $p_{\alpha_{2,1}} < .001$) and 3 h ($\alpha_{2,3} = 0.037$, $SE_{\alpha_{2,3}} = 0.016$, $p_{\alpha_{2,3}} = .020$) later. The cross-regressive relationships suggested that nervousness significantly predicted heart rate 8 h later ($\gamma_{1,2,8} = 0.055$, $SE_{\gamma_{1,2,8}} = 0.020$, $p_{\gamma_{1,2,8}} = .007$). Interestingly, nervousness significantly negatively predicted heart rate 12 h later ($\gamma_{1,2,12} = -0.052$, $SE_{\gamma_{1,2,12}} = 0.024$, $p_{\gamma_{1,2,12}} = .030$). Next, we compared the magnitude of the effect of nervousness on heart rate 8 h later compared to the effect of 12 h later by removing the lagged effects alone and examining the differences in model fit by means of a likelihood ratio test. The change in model fit ($-2 \times \log$ likelihood) was greater when the cross-regressive coefficient of 8 h was removed ($-2 \text{ LL} = 7.29$ with 1 df), compared to that observed when the coefficient for 12 h later was removed ($-2 \text{ LL} = 4.68$ with 1 df). These results indicate

DTVEM Stage 1

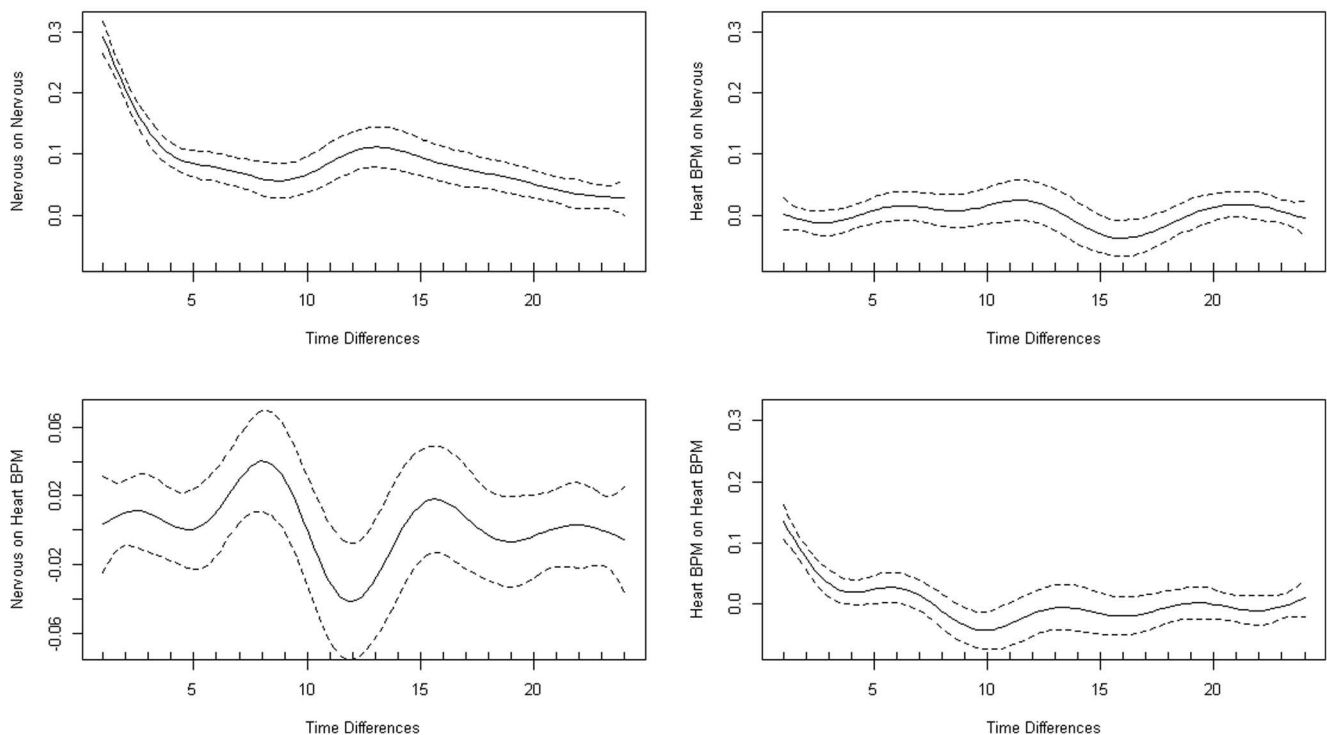


Fig. 7 The results depict the exploratory stage of DTVEM for nervousness and heart rate, predicting themselves and each other. The solid lines depict the autoregressive and cross-regressive coefficients

over all possible time differences. The dashed lines depict the confidence bands of the regression coefficients.

that the effect of nervousness on heart rate was stronger 8 h later than 12 h later. Notably, the model was tested for stationarity, and was stationary (based on all the values of the modulus of the roots of the determinant of the identity matrix minus the transition matrix being greater than one).

In sum, we did not find a bidirectional relationship between nervousness and heart rate, but rather, that nervousness unidirectionally drove the degree of heart rate 8 and 12 h later. A hormonal response may explain the significantly positive relationship followed by a significantly negative relationship between nervousness and later heart rate. Specifically, nervousness predicts the release of norepinephrine, and norepinephrine has a positive association with heart rate that quickly dissipates over approximately the same number of hours (Mergler & Valciukas, 1998). This illustrates the utility of DTVEM in providing a convenient automated framework to estimate unknown lag structures.

Simulation study

A simulation study was performed to test the performance of DTVEM at detecting optimal lag structures in processes that unfold over diverse time scales. Our goal was to consider sample size and design conditions that mirrored empirical processes of varying complexity (e.g., involving processes unfolding over a single vs. multiple time scales, with univariate vs. multivariate data, and complete vs. missing data).

First, we chose one of the most common AR structures in the behavioral sciences, an AR model of order 1 with two sample size configurations: (1a) $T = 100$, $N = 1$, a configuration commonly utilized in the time series literature to evaluate finite-sample performance of statistical approaches; and (1b) $T = 14$, $N = 100$, as commonly seen in many ILD studies in psychology (Armstrong et al., 2010; Beidel, Neal, & Lederer, 1991; Roche et al., 2016; Steger & Frazier, 2005). Second, we simulated a complex AR structure with sharp shifts using a seasonal autoregressive integrated moving average (SARIMA; with $N = 100$, $T = 14$) (Box & Jenkins, 1976; Box et al., 2013; Brockwell & Davis, 2002; Reinsel, 2003).

Third, we wanted to test a typical multivariate case ($N = 159$, $T = 170$) that mirrored the irregular measurement intervals and complex lag structure and found in the empirical example. Our goal was to evaluate the performance of the DTVEM in recovering complex group-based, multivariate, lag structure using multiple-subject data. This simulation study served to extend the results from previous simulation studies in which researchers compared inferential results obtained from fitting discrete-time models to unequally spaced assessments that were binned or blocked together at

prespecified windows (e.g., hourly, every 4 h) with missingness inserted, in which appropriate, to create a set of equally spaced data. These designs tend to be highly prevalent in ILDs (Ebner-Priemer, Eid, Kleindienst, Stabenow, & Trull, 2009), and blocking irregularly spaced data at equally spaced time windows with missingness inserted in which appropriate is a relatively common way of handling irregular spacing in ILDs (Silvia, Kwapil, Walsh, & Myin-Germeys, 2014).

Researchers (de Haan-Rietdijk, Voelkle, Keijsers, & Hamaker, 2017) have found in a previous simulation study that applying this method of data blocking to single-subject, equally spaced time series blocked over successive windows yielded inferential results that closely approximated those obtained from continuous-time models in situations involving a lag-1 structure. However, this method of blocking was not as successful in approximating continuous-time models in other scenarios involving continuous-time models that capture higher-order changes (and hence higher-order lag structure in their discrete-time counterparts), particularly when the time intervals used to define successive lags are large relative to those used in data generation (de Haan-Rietdijk et al., 2017). In other words, de Haan-Rietdijk et al. considered scenarios in which the time intervals used in data generation might be different from those used to define the lags (e.g., measurement occasions that were separated by an interval of $\Delta t = 2$, as opposed to those separated by $\Delta t = 1$, were specified as one lag apart). Here we generated data using discrete-time models in which Δt was set to 1.0 both in the data generation and model-fitting process. Our goal is to determine, under situations with correctly specified Δt , the performance of the DTVEM when applied to multisubject data with an unknown lag structure, a large amount of missing data (72%), and small signal-to-noise ratios.⁶

We conducted 500 Monte Carlo (MC) simulations for each of the three primary modeling variations considered. These variations include (1a) an AR(1) model with $N = 100$, $T = 14$; (1b) an AR(1) model with $n = 1$, $T = 100$; (2) a SARIMA model with $N = 100$, $T = 14$; and (3) a multivariate case with $N = 159$, $T = 170$, and 72% missing data. The ratio between the simulated values and the total variance (i.e., the signal-to-noise ratio) ranged from 0.10 to 0.60 (the process noise was set to 2.25 for all models). The DTVEM was applied without any a priori knowledge about the lag structure and allowed to go through iterative cycles of the lag exploration and confirmatory VAR(p) model-fitting stage until no significant lags were identified in the additional exploratory stages of DTVEM.

⁶ Note that signal-to-noise ratios were estimated empirically by simulating $N = 1$ and $t = 1,000,000$, and estimating the proportion of variation explained by including the true lags (R^2), then calculated using the signal-to-noise ratio, $R^2/(1 - R^2)$ (Cohen, 1988; Kelley, 2007; Muirhead, 1985).

Simulation 1: AR(1) simulation

The first simulation was based on an AR process of order 1 [i.e., AR(1) process] with univariate equidistant time points and complete cases. Time series data were generated using one of two possible sample size configurations: (Simulation 1a) $N = 100$ persons over $T = 14$ time points, and (Simulation 1b) $n = 1$ person over $T = 100$ time points, using the simulation model:

$$y_{i,t} = .5*(y_{i,t-1}) + \varepsilon_{i,t}$$

In this formula, y is the outcome variable for person i at time t , $\varepsilon_{i,t}$ represents the process noise for person i and time t , assumed to be normally distributed. The process noise was set to be 2.25, and, as such, the signal-to-noise ratio was 0.34.

Simulation 2: Seasonal autoregressive integrated moving average [SARIMA(1,0,0) $\times$ (1,0,0)₇] simulation

Next, we tested DTVEM using a process that unfolds over multiple time scales by pairing an AR model with a seasonal component (i.e. SARIMA, e.g., a day-to-day process embedded in a weekly cycle; hour-by-hour fluctuations with diurnal rhythms embedded) (Box & Jenkins, 1976; Brockwell & Davis, 2002; Reinsel, 2003). In this case, we included a seasonal component with a period of 7, formulated such that the current cycle was affected by the immediately preceding cycle. That is, there was an AR(1) component for this seasonal component, such that the current level of the process was affected by the level of the process seven lags ago. Beyond this seasonal component, the process was hypothesized to also show lag-1 AR dynamics, yielding a SARIMA(1,0,0) $\times$ (1,0,0)₇ model. Our specific choice of AR weights for the seasonal and non-seasonal components yielded the following combined model:

$$y_{i,t} = .5*(y_{i,t-1}) + .4*(y_{i,t-7}) - .2*(y_{i,t-8}) + \varepsilon_{i,t}$$

In this formula, y is the outcome variable for person i at time t , and $\varepsilon_{i,t}$ represents the process noise for person i and time t , assumed to be normally distributed. Note that the SARIMA has complex AR dynamics, as the process has a significantly positive lag at 7 and has a significantly negative lag one lag later.⁷ The signal-to-noise ratio for this model was 0.60.

⁷ Specifically, to understand the full dynamics of multiplicative SARIMA models such as the SARIMA(1,0,0) $\times$ (1,0,0)₇ model considered here, one can first express the model using the lag operator and expand the lag polynomials as follows. Specifically, let L denote the lag operator, such that $L(x_t) = x_{t-1}$ and $L^p(x_t) = x_{t-p}$. The SARIMA(1,0,0) $\times$ (1,0,0)₇ model considered here can be expressed using the lag operator as (Box & Jenkins, 1976; Brockwell & Davis, 2002; Hamilton, 1994; Reinsel, 2003):

$$(1 - \alpha_{17}L^7)(1 - \alpha_{11}L)y_{i,t} = \zeta_{y,i,t}(1 - \alpha_{11}L - \alpha_{17}L^7 + \alpha_{17}\alpha_{11}L^8)y_{i,t} = \zeta_{y,i,t}y_{i,t} = \alpha_{11}y_{i,t-1} + \alpha_{17}y_{i,t-7} + \alpha_{18}y_{i,t-8} + \zeta_{y,i,t} \text{ where we set } \alpha_{11} = 0.5, \alpha_{17} = 0.4, \text{ and } \alpha_{18} = -\alpha_{17}\alpha_{11} = -0.2.$$

Simulation 3: Multivariate time series simulation with 72% missingness

Thus far, all simulations of DTVEM have been univariate with no missingness. As such, we wanted to test the validity of DTVEM with multivariate data that more closely resembled typical-use scenarios with ILD. Inspired by the empirical example, we created a simulation condition using the same number of persons, with the same pattern of missingness, and the same lag structure to generate a more realistic scenario. Thus, the multivariate simulations were generated through two simulation equations:

$$\begin{aligned} y_{1,i,t} &= .3*(y_{1,i,t-1}) + .1*(y_{1,i,t-3}) + .1*(y_{1,i,t-4}) + .1*(y_{1,i,t-12}) + .1*(y_{1,i,t-14}) + \varepsilon_{y_1,i,t} \\ y_{2,i,t} &= .1*(y_{2,i,t-1}) + .1*(y_{2,i,t-3}) + .2*(y_{2,i,t-8}) - .2*(y_{1,i,t-12}) + \varepsilon_{y_2,i,t} \end{aligned}$$

In this formula, two time series were simulated together. In these cases, y_1 and y_2 represented the two time-series outcomes for person i at time t , and $\varepsilon_{y_1,i,t}$ and $\varepsilon_{y_2,i,t}$ represented the process noise for person i and time t for y_1 and y_2 , respectively, which was assumed to be normally distributed with zero means. The process noise variances for both time series were set to 2.25, as in the other models. Note that there were five separate autoregressive lags for y_1 , at 1, 3, 4, 12, and 14, and two autoregressive lags for y_2 , at 1 and 3. Additionally, the outcome variable y_2 was also influenced by y_1 at lags of 8 and 12. Thus, we increased the size of the lag coefficients to increase the signal to noise ratio within the data (between absolute value strengths of 0.1 to 0.3). Nevertheless, the signal-to-noise ratios were still quite small, at 0.23 for y_1 and 0.10 for y_2 .

Model evaluation

All DTVEM estimates were examined through the following metrics: (1) the number of auto- and cross-regressive parameters estimated, (2) the mean estimated values across simulations (i.e., the mean $\hat{\theta}$ values); (3) the root mean square error

(RMSE), which is $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)^2}$; (4) bias, which is defined by $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)$; (5) the mean SE across MC runs, as compared to the standard deviation of the estimates across MC runs; (6) the mean relative deviance of the SE (RDSE), which is defined by $\frac{\text{Mean SE} - \text{SD of } \hat{\theta} \text{ across MC runs}}{\text{SD of } \hat{\theta} \text{ across MC runs}}$; (7) coverage, which is the proportion of 95% confidence intervals whose values include the simulated θ across MC simulations; (8) power, defined as the proportion of parameter estimates for truly non-zero population parameters from the confirmatory stage across the MC runs whose 95% confidence intervals do not contain zero; (9) the average Type I error rate, which is defined as the proportion of MC replications for which the 95% confidence intervals incorrectly identified any one spurious lag as

statistically significant, (as averaged over all possible spurious lags considered); and (10) the total Type I error rate, which is defined as the proportion of MC replications for which the 95% confidence intervals incorrectly detected one or more spurious lags. Note that we considered a maximum lag of 10 for each simulation. We compared the final confirmatory estimates from DTVEM to a state-space model with lagged structures that coincided with the data generation models. That is, to provide a benchmark against which the DTVEM estimates were compared, we obtained ML estimates under the “ideal” scenario that a priori knowledge concerning the lag structures was already available—a scenario that is often not fulfilled in practice. Estimation was performed using OpenMx (Neale et al., 2016) through state-space specification of a VAR(p) model and was subjected to the same evaluation criteria as the DTVEM.

Results

Simulation 1: AR(1) results with no missingness

For AR(1) with one person with 100 time points and 100 persons with 14 time points, the results were grouped together because they were very similar. The exploratory stage passed on an average of three autoregressive parameters to be estimated in the confirmatory stage (see Table 1). A summary of the findings is depicted in Table 2, and comparison results from using confirmatory state-space models with correctly specified lag structures are shown in Table 3. Power estimates from the DTVEM suggested that the powers for detecting the truly nonzero parameters as being significantly different from zero were 99.8% and 100% for the two sample size configurations considered (with $N = 1$, $T = 100$, and $N = 100$, $T = 14$, respectively), as compared to 100% power for both simulation conditions in the confirmatory models. Both DTVEM and the confirmatory models also showed 100% power to detect the truly nonzero AR(1) parameter and process noise variance across both simulations. For the $N = 1$ and $T = 100$ condition, we considered up to lag 10 and found an average Type I error rate of 1.6% and a total Type I error rate of 11.8%. Likewise, for $N = 100$ and $T = 14$, the average Type I error rate was 1.6%, and the total Type I error rate was 8.6%. For both simulation conditions, the biases and RMSEs of the point estimates were very small and closely paralleled the biases and RMSEs from the confirmatory models. For both conditions, the standard error estimates also closely mirrored the “true” or empirical standard errors, and coverage rates for all time series parameters were close to the 95% nominal level. The results of this condition suggest that the

combined DTVEM procedures led to excellent recovery of the lag structure and corresponding estimates in the absence of a priori knowledge concerning the lag structure. Unlike standard exploratory approaches such as PACF plots, DTVEM worked for both single-subject time series of moderate length as well as multiple-subject replicated (but short) time series.

Simulation 2: SARIMA(1,0,0) $\times$ (1,0,0)₇ results with 14 time points with no missingness

The DTVEM estimates and SE were similar to those from previous AR(1) conditions. On the basis of variable selection in the exploratory stage, an average of four autoregressive parameters were estimated at the confirmatory stage (see Table 1). The process noise variance estimate for $\zeta_{y, i, j}$ was characterized by slight increases in RMSE and bias, as compared to the AR(1) conditions, reflecting the increased difficulties in estimating this parameter with more complex lag structures and limited time series length. The power was slightly decreased for part of the estimates as compared to prior models, with DTVEM showing 100% power to detect the nonzero lag values of 1 and 7, and 98.4% power in detecting the lag of 8. The high power across conditions is noteworthy, given the small value of the simulated lag-8 coefficient and the limited replications of data spanning eight lags apart. For the seven lags considered (we considered up to lag 10, and thus lags 2–6, 9, and 10 were truly zero), the average Type I error rate was 1.2%, and the total Type I error rate was 8.6%. As with the previous results, the estimates and standard errors were satisfactory and closely mirrored those from the correctly specified confirmatory state-space model (see Table 3).

Simulation 3: Multivariate time series results with 72% missing data and a complex lag structure

The DTVEM point and SE estimates in this model showed trends similar to those from the previous models (i.e., satisfactory point and SE estimates that closely mirrored the correctly specified confirmatory model; see Tables 4 and 5). On the basis of variable selection at the exploratory stage, an average of 27 auto- and cross-regressive parameters were estimated at the confirmatory stage (see Table 1). This means that 55% of the potential lags were eliminated from the exploratory stage (relative to only 15% of potential lags that were simulated to be nonzero in the population). The coverage rates for all other parameters were close to the .95 nominal level. Power and the Type I error rate were similar to the values from prior models, with DTVEM showing approximately 100% power to detect the lag-1, lag-3, lag-4, lag-12, and lag-14 autoregressive parameters of y_1 , as well as power in

Table 1 Numbers of parameters estimated on the basis of variable selection at the DTVEM exploratory stage

Simulation	<i>T</i>	% <i>M</i>	# of Nonzero Regressive Parameters	<i>M</i> Number of Regressive Parameters Estimated	<i>SD</i> # of Regressive Parameters Estimated	5th Quantile of Number of Regressive Parameter Estimates	95th Quantile of Number of Regressive Parameter Estimates
AR(1), <i>N</i> = 100	100	0	1	3.288	1.926	1	7
AR(1), <i>N</i> = 14	14	0	1	3.948	1.767	1	7
SARIMA	14	0	3	4.092	0.939	3	6
Multivariate	170	72	9	27.484	4.473	21	36

This table predicts the number of auto- and cross-regressive parameters estimated on the basis of significant lags extracted from the exploratory stage of DTVEM. In all simulations, DTVEM extracted a larger number of parameters than non-zero parameters on average. Nevertheless, these simulations demonstrated, the exploratory stage of DTVEM reduced the number of parameters estimated in confirmatory models substantially across every simulation in every case.

detecting the lag-8 and lag-12 cross-regressive parameters of y_1 on y_2 and the lag-1 autoregressive parameter of y_1 . Nevertheless, the power to detect the lag-3 autoregressive parameter for y_2 was 61%, likely resulting from a very low signal-to-noise ratio. Regarding the AR relationships with y_1 as an outcome, the average Type I error rate was 2.8%, and the total Type I error rate was 10.0%. For the AR relationships with y_2 as an outcome, the average Type I error rate was 2.1%, and the total Type I error rate was 8.3%. The average Type I error rate for the cross-regressive relationship of y_2 on y_1 was 1.1%, and the total Type I error rate was 8.3%. For y_1 on y_2 , the average Type I error rate was 3.1%, and the total Type I error rate was 15.4%. Overall, these results suggest a slight increase in difficulty for DTVEM to simultaneously identify the correct lag structures when multiple constructs were involved

with a complex lag structure, and there was a large proportion of missing data. Nevertheless, the high power for detecting most of the truly nonzero lag coefficients, with the relatively low average and total Type I error rates of the DTVEM in such a challenging scenario, provide some reassurance of the proposed approach's utility in handling complex lag structures in other similar empirical settings.

Summary of simulation results

The simulations suggest that DTVEM has high exploratory power and low Type I error rates across both univariate and bivariate models with no a priori information concerning the lag structures. DTVEM's point estimates were accurate across the simulation conditions. The *SE* estimates were unbiased for all conditions, either without missing data or with 72% of the

Table 2 Summary of DTVEM confirmatory-stage estimates for complete data

Simulation	<i>T</i>	Term	Simulated θ	Mean $\hat{\theta}$	RMSE	Bias	<i>SD</i> of $\hat{\theta}$	Mean of <i>SE</i>	RDSE	Coverage of 95% CI	Power
AR(1), <i>N</i> = 1	100	AR(1)	0.500	0.449	0.051	-0.051	0.092	0.090	-0.031	0.906	0.998
AR(1), <i>N</i> = 1	100	$\zeta_{y, i, j}$	2.250	2.152	0.098	-0.098	0.315	0.305	-0.033	0.882	1.000
AR(1), <i>N</i> = 100	14	AR(1)	0.500	0.499	0.001	-0.001	0.024	0.024	0.015	0.964	1.000
AR(1), <i>N</i> = 100	14	$\zeta_{y, i, j}$	2.250	2.265	0.015	0.015	0.091	0.086	-0.060	0.938	1.000
SARIMA	14	AR(1)	0.500	0.510	0.010	0.010	0.025	0.025	-0.021	0.918	1.000
SARIMA	14	AR(7)	0.400	0.393	0.007	-0.007	0.039	0.037	-0.039	0.936	1.000
SARIMA	14	AR(8)	-0.200	-0.209	0.009	-0.009	0.038	0.041	0.093	0.956	0.832
SARIMA	14	$\zeta_{y, i, j}$	2.250	2.407	0.157	0.157	0.109	0.097	-0.108	0.634	1.000

This is the summary of the DTVEM confirmatory stage estimates. Simulation refers to the way in which the data were simulated. *T* refers to the total number of time points for each person. AR refers to whether the simulated model term of interest was autoregressive. The symbol Δt refers to the time difference in which the model was simulated. These results depict (1) simulated θ (the ways in which the data were simulated); (2) the mean $\hat{\theta}$ values; (3) the root mean square error (RMSE), which is $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)^2}$; (4) bias, which is defined by $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)$; (5) the standard deviation of the $\hat{\theta}$ across Monte Carlo runs; (6) the mean standard error (*SE*) across Monte Carlo runs; (7) the mean relative deviance of the *SE* (RDSE), which is defined by $\frac{\text{Mean } SE - SD \text{ of } \hat{\theta}}{SD \text{ of } \hat{\theta}}$; and (8) power, defined as the proportion of parameter estimates from the confirmatory stage across the MC runs whose 95% confidence intervals do not contain zero.

Table 3 Summary of confirmatory state-space estimates for complete data

Simulation	T	Term	Simulated θ	Mean $\hat{\theta}$	RMSE	Bias	SD of $\hat{\theta}$	Mean of SE	RDSE	Coverage of 95% CI	Power
AR(1), $N = 1$	100	AR(1)	0.500	0.496	0.004	-0.004	0.086	0.087	0.011	0.954	1.000
AR(1), $N = 1$	100	$\zeta_{y, i, j}$	2.250	2.242	0.008	-0.008	0.319	0.317	-0.005	0.926	1.000
AR(1), $N = 100$	14	AR(1)	0.500	0.501	0.001	0.001	0.024	0.024	0.021	0.958	1.000
AR(1), $N = 100$	14	$\zeta_{y, i, j}$	2.250	2.271	0.021	0.021	0.092	0.087	-0.055	0.936	1.000
SARIMA	14	AR(1)	0.500	0.514	0.014	0.014	0.024	0.025	0.019	0.902	1.000
SARIMA	14	AR(7)	0.400	0.399	0.001	-0.001	0.035	0.035	-0.007	0.948	1.000
SARIMA	14	AR(8)	-0.200	-0.210	0.010	-0.010	0.037	0.040	0.078	0.956	1.000
SARIMA	14	$\zeta_{y, i, j}$	2.250	2.389	0.139	0.139	0.098	0.097	-0.011	0.712	1.000

This is the summary of the confirmatory state-space estimates across each of the simulations with complete data. Simulation refers to the way in which the data were simulated. T refers to the total number of time points for each person. AR refers to whether the simulated model term of interest was autoregressive. The symbol: Δt refers to the time difference in which the model was simulated. These results depict (1) simulated θ (the ways in which the data were simulated); (2) the mean $\hat{\theta}$ values; (3) the root mean square error (RMSE), which is $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)^2}$; (4) bias, which is defined by $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)$; (5) the standard deviation of the $\hat{\theta}$ across Monte Carlo runs; (6) the mean standard error (SE) across Monte Carlo runs; (7) the mean relative deviance of the SE (RDSE), which is defined by $\frac{\text{Mean } SE - SD \text{ of } \hat{\theta}}{SD \text{ of } \hat{\theta}}$; (8) coverage, which is the proportion of simulated values that fall within the 95% confidence intervals across Monte Carlo simulations; and (9) power, defined as the proportion of parameter estimates across the MC runs whose 95% confidence intervals do not contain zero.

data missing. The computational time needed for the R DVTEM program to manipulate and perform the analyses was under 2 min for each of the present simulations, except for the multivariate simulation, which took approximately 2 h

to run [note that this was primarily due to an expansion of the Hessian matrix in the state-space models, in which the estimation process for the number of subjects multiplied by the number of potential lags squared = $(159 \times 30)^2 = 26,010,000$].

Table 4 Summary of DTVEM confirmatory-stage estimates with missing data

Simulation	T	%M	Term	Simulated θ	Mean $\hat{\theta}$	RMSE	Bias	SD of $\hat{\theta}$	Mean of SE	RDSE	Coverage of 95% CI	Power
Multivariate	170	72	AR(1)	0.300	0.300	0.013	0.000	0.013	0.011	-0.157	0.924	1.000
Multivariate	170	72	AR(3)	0.100	0.102	0.015	-0.012	0.014	0.013	-0.076	0.911	0.948
Multivariate	170	72	AR(4)	0.100	0.102	0.019	-0.012	0.019	0.013	-0.291	0.911	0.988
Multivariate	170	72	AR(12)	0.100	0.100	0.015	0.000	0.015	0.014	-0.074	0.935	0.986
Multivariate	170	72	AR(14)	0.100	0.099	0.015	0.001	0.015	0.014	-0.101	0.936	1.000
Multivariate	170	72	AR(1)	0.100	0.101	0.013	-0.001	0.013	0.012	-0.048	0.944	1.000
Multivariate	170	72	AR(3)	0.100	0.098	0.013	0.002	0.013	0.013	0.019	0.950	0.606
Multivariate	170	72	CR(8)	0.200	0.202	0.017	-0.002	0.017	0.015	-0.110	0.930	1.000
Multivariate	170	72	CR(12)	-0.200	-0.201	0.018	0.001	0.018	0.015	-0.142	0.950	1.000
Multivariate	170	72	$\zeta_{y_1, i, j}$	2.250	2.243	0.033	0.007	0.032	0.031	-0.032	0.938	1.000
Multivariate	170	72	$\zeta_{y_2, i, j}$	2.250	2.247	0.037	0.003	0.038	0.035	-0.080	0.960	1.000

This is the summary of the DTVEM estimates across each of the missing data simulations. All estimates presented in this table are from the DTVEM confirmatory stage. Simulation refers to the way in which the data were simulated. T refers to the number of time points, %M refers to the percentage of missing data which was missing at random and was different for each person, and AR/CR refers to whether the simulated model term of interest was autoregressive (AR) or cross-regressive (CR). The symbol Δt refers to the time difference in which the model was simulated. These results depict (1) simulated θ (the ways in which the data were simulated); (2) the mean $\hat{\theta}$ values; (3) the root mean square error (RMSE), which is $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)^2}$; (4) bias, which is defined by $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)$; (5) the standard deviation of $\hat{\theta}$ across Monte Carlo runs; (6) the mean standard error (SE) across Monte Carlo runs; (7) the mean relative deviance of the SE (RDSE), which is defined by $\frac{\text{Mean } SE - SD \text{ of } \hat{\theta}}{SD \text{ of } \hat{\theta}}$; and (8) power, defined as the proportion of parameter estimates from the confirmatory stage across the MC runs whose 95% confidence intervals do not contain zero.

Table 5 Summary of confirmatory state-space estimates with missing data

Simulation	<i>T</i>	%M	Term	Simulated θ	Mean $\hat{\theta}$	RMSE	Bias	<i>SD</i> of $\hat{\theta}$	Mean of <i>SE</i>	RDSE	Coverage of 95% CI	Power
Multivariate	170	72	AR(1)	0.300	0.299	0.011	0.001	0.011	0.011	0.015	0.952	1.000
Multivariate	170	72	AR(3)	0.100	0.100	0.013	0.000	0.013	0.013	− 0.013	0.950	1.000
Multivariate	170	72	AR(4)	0.100	0.101	0.014	− 0.001	0.014	0.013	− 0.032	0.950	1.000
Multivariate	170	72	AR(12)	0.100	0.100	0.014	0.000	0.014	0.013	− 0.028	0.948	1.000
Multivariate	170	72	AR(14)	0.100	0.100	0.014	0.000	0.014	0.014	− 0.013	0.952	1.000
Multivariate	170	72	AR(1)	0.100	0.099	0.012	0.001	0.012	0.012	0.016	0.952	1.000
Multivariate	170	72	AR(3)	0.100	0.100	0.012	0.000	0.012	0.013	0.065	0.960	1.000
Multivariate	170	72	CR(8)	0.200	0.202	0.015	− 0.002	0.015	0.014	− 0.013	0.954	1.000
Multivariate	170	72	CR(12)	− 0.200	− 0.200	0.015	0.000	0.015	0.015	− 0.054	0.936	1.000
Multivariate	170	72	$\zeta_{y_1,i,j}$	2.250	2.244	0.032	0.006	0.032	0.031	− 0.026	0.938	1.000
Multivariate	170	72	$\zeta_{y_2,i,j}$	2.250	2.240	0.036	0.010	0.035	0.034	− 0.022	0.944	1.000

This is the summary of the confirmatory estimates across each of the missing data simulations. Simulation refers to the way in which the data were simulated. *T* refers to the total number of time points for each person, %M refers to the percentage of missing data which was missing at random and was different for each person, and AR/CR refers to whether the simulated model term of interest was autoregressive (AR) or cross-regressive (CR). The symbol Δt refers to the time difference in which the model was simulated. These results depict (1) simulated θ (the ways in which the data were simulated); (2) the mean $\hat{\theta}$ values; (3) the root mean square error (RMSE), which is $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)^2}$; (4) bias, which is defined by $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - \theta)$; (5) the standard deviation of $\hat{\theta}$ across Monte Carlo runs; (6) the mean standard error (*SE*) across Monte Carlo runs; (7) the mean relative deviance of the *SE* (RDSE), which is defined by $\frac{\text{Mean } SE - SD \text{ of } \hat{\theta}}{SD \text{ of } \hat{\theta}}$; (8) coverage, which is the proportion of simulated values that fall within the 95% confidence intervals across Monte Carlo simulations; and (9) power, defined as the proportion of parameter estimates across the MC runs whose 95% confidence intervals do not contain zero.

Discussion

This article has described a convenient user-accessible exploratory and confirmatory tool, DTVEM, which allows users to model time lags in the state-space framework. Using a series of simulation examples, we found that DTVEM had high power in the detection of optimal lag structures, low Type I error rates, accurate point estimates, and accurate *SEs* as long as sufficient measurement occasions were available to recover higher-order lags. The simulation results support the utility of DTVEM as an exploratory tool over the partial autocorrelation function with multisubject data. The simulation results are particularly notable given the high dimensionality of the lags considered. For example, the multivariate simulation across 15 lags with two dependent variables is analogous to 30 independent predictors across each of the two outcomes (i.e., 60 potential regression estimates). Thus, DTVEM demonstrated high performance across a diverse number of conditions with high dimensionality, low signal to noise ratio, and high missingness.

Within the behavioral sciences, DTVEM can be used to answer a variety of research questions involving different ILDs such as daily diary studies, ecological momentary assessments, physiological, and neurological data collection. Examples of research questions the DTVEM can facilitate include: (1) Does one construct predict itself or another construct at a later time? (this could be detected through auto- and cross-regressive lags); (2) When does one construct

maximally predict other constructs? (this could be detected through the consideration of many distinct lags beyond lag 1); and (3) Are oscillatory or other cyclic patterns present in my data? Such patterns are typically manifested as oscillatory patterns in the lag coefficients, such as significant positive lag coefficients followed by significant negative lag coefficients, or vice versa. Weekly cycles may be revealed through the inspection of statistically significant coefficient at lag 7 with daily data, whereas a yearly cycle can be capture as significant lag-12 coefficient with monthly data. Taken together, DTVEM has a substantially broad application and has considerable utility for a number of research paradigms. For example, DTVEM would allow (1) psychopathology researchers to uncover maintenance or short-term risk factors within psychopathology (i.e., where within-person variation in one construct is largely driven by within-person variation in another construct), (2) personality and social psychology researchers to explore dynamics in behavior across several time periods, or (3) researchers to explore how and when developmental processes occur (e.g., how and when does a mother's behavior influence a child's behavior, and vice-versa).

The DTVEM comprises relatively well-known (specifically, GAMM and state-space estimation) subroutines and, as such, may not be seen as a novel tool. However, DTVEM and the results highlighted in this article make several novel contributions to the field. First, DTVEM can be seen as a utility function that features novel integration and application of some of the key strengths of the GAMM and state-space

modeling frameworks toward the issue of lag detection. Second, this novel combination of GAMM and state-space subroutines results in high detection of the simulated lag structure with satisfactory point estimates, reasonable standard error estimates, and relatively high power. Additionally, this article builds on prior findings that evaluated the feasibility of drawing inferential conclusions concerning the dynamics of a system when discrete-time models are fitted to irregularly spaced data that are binned at equally spaced time blocks (de Haan-Rietdijk et al., 2017). Our stimulation results extend those findings in three notable ways: (1) The results suggest that this method of binning irregularly spaced data—which inevitably leads to large proportions of missingness—yielded reasonable inferential results with DTVEM when the lag order used in model fitting mirrored the true lag order in data generation defined with $\Delta t = 1$; (2) they provide an assessment of the extent to which variable selection tools in the regression framework such as GAMM can be used to uncover unknown but potentially complex lag structures; and (3) they reveal the quality of the point and standard error estimates, as well as the strengths and limitations of the proposed approach when used to estimate VAR models of complex lag structures with little a priori knowledge on the lag structures of the data. Finally, by offering this tool within a convenient user-accessible format, it allows users who are not overly familiar with either GAMM or state-space modeling frameworks to run these models in a more accessible and highly automated fashion.

One of the most important aspects of the DTVEM is its ability to easily help users visualize and explore possible changes in the dependent variables as linear as well as nonlinear functions of time, time differences, and other variables implicated in the system's dynamics. For instance, although many systems that conform to a simplex structure (Guttman & Guttman, 1965) show reduced autocorrelations between occasions that are closely spaced in time than those that are further apart—thereby manifesting a monotonic decrease in autocorrelations with an increase in time differences (or time lags)—abundant examples exist that point to the contrary. Consider the results of the empirical example, in which nervousness predicted heart rate 8 and 12 h later. In this case, the partial cross-correlational structure of the data changes as a nonmonotonic function of time differences as the cross-regression effect of nervousness on heart rate does not degrade monotonically as time elapses—as would be predicted by a simplex data structure—but rather shows abrupt jumps at particular lags. Processes that are characterized by only higher-order lags are also commonly seen in seasonal (e.g., diurnal, weekly, monthly, annually) processes that only manifest significant associations with prior levels at specific lags (e.g., a strictly weekly process may only show a significant lag at a time interval of 7, but no other, lower-order lags). The auto- and cross-correlational structures of empirical data may also show further dependencies on other variables, such as time

(Chow et al., 2011), spatial region (Crainiceanu, Ruppert, Carroll, Joshi, & Goodner, 2007), and other intra- and inter-individual difference characteristics (e.g., Chow & Zhang, 2013).

Other approaches have been previously used to select and model optimal lag structures. For example, Ho, Shumway, and Ombao (2006) used the Bayesian information criterion (BIC) to perform model selection in VAR(1) model. Such an information criterion-based approach requires that a researcher fits all possible candidate models and compares their BIC values before selecting the final model. Naturally, the number of candidate models considered is typically small (e.g., Ho et al., 2006, only considered up to one lag) to render the model comparison process computationally feasible. As an example, consider the number of models required for two variables with 15 time lags. This would require fitting 60 separate models (there are 60 possible lag coefficients), followed by a possibility of exploring all potentially significant lags. This could in practice require a total of $60! = 8.32 \times 10^{81}$ models to be fit, if selection was based on the BIC. Higher-order lags might be ignored and as we have illustrated, such omission can lead to fundamentally incorrect inferential conclusion. In addition, this approach alone is only practically feasible in particularly small-use scenarios (i.e., both small number of variables and a small number of lags).

Admittedly, DTVEM is designed to be a preliminary tool. As an example, if lags of 3 and 12 were identified by means of DTVEM in a study involving hourly measurements, this then raises the question of whether the process of interest should be modeled as an AR process with nonzero coefficients at lags of 3 and 12 [an AR(3,12) process] or an AR process with nonzero coefficients at lags of 1 and 4 [an AR(1,4) process], with each lag spaced 4 h apart. Ultimately, DTVEM, as it stands, is meant to serve as an accessible exploratory tool—a first step toward identifying the proper lag structures for a specific class of models. The final choice and model formulation still depend on characteristics of the constructs under study and the change characteristics expected of them.

In addition to the strengths of the DTVEM, it is important to note that the approach also has limitations and other unresolved issues. First, DTVEM operates only to detect the optimal lag structure in discrete-time VAR(p) models. In the future, DTVEM could be modified to allow for continuous-time modeling variations (e.g., Oud, 2004) or other hybrid continuous-discrete variations (Bar-Shalom, Li, & Kirubarajan, 2001). However, just as lag selection is a central issue in fitting discrete-time VAR models, users interested in fitting continuous-time models have to select the order of their models (i.e., the highest-order changes, or derivatives, involved in the model), which usually take on the form of ordinary differential equations (ODEs) and stochastic differential equations (SDEs). Thus, continuous-time models are not immune from the issue of lag selection, although they do have

some advantages in that the results from continuous-time models are not affected by the measurement time intervals used in a particular study, whereas time intervals and lags can be confounded in discrete-time models. Nevertheless, dominant continuous-time models have their own limitations and computational difficulties. For example, the exact discrete-time approach advocated by Voelkle, Oud, Davidov, and Schmidt (2012) requires one to work with the solutions of SDEs. Not all SDEs have known analytic forms, and even when they do, the solutions can take on very different forms depending on the order of SDEs, and often they require the specification of complex constraints in the model-fitting process. The results from models that include high-order derivatives (e.g., beyond rates of change and changes in the rates of change) may also be difficult to interpret substantively. In contrast, there is a well-established statistical and econometric literature on how to diagnose, explore, estimate, and interpret results from discrete-time models such as SARIMA models. Finally, some special cases of discrete-time models cannot be obtained in the continuous-time framework [e.g., AR(1) process with an AR(1) weight of -1]. We limited ourselves to lag detection in a discrete-time framework in the present article. Future studies should extend this line of work to continuous-time models.

Additionally, the present simulation studies did not consider person-specific differences in dynamics or lag structures. Given this, future extensions of the DTVEM model may be able to include random effects within the GAMM framework to model person-specific differences such as individual differences in intercept, as well as the issue of lag identification in discrete data. Although the number of parameters estimated were large within simulations (i.e., 60 lag parameters in Simulation 3), the present simulation studies were limited to bivariate VAR models with two dependent variables. As such, our conclusions are limited in generalizability to particular kinds of VAR models. Future work should add simulation by increasing the number of independent and dependent variables considered, especially given that intensive longitudinal datasets often are collected with a large number of variables. Doing this would be especially important given the increased popularity of network psychometric models, and extensions of these models to incorporate VAR-based lag structures within network analysis (Epskamp et al., 2018; Marsman et al., 2018). Moreover, as the inclusion of latent variables is currently not feasible within the GAMM framework, we have restricted our attention to a VAR(p)—namely, a manifest variable framework. An alternative framework focuses on evaluations of lag structures in models involving latent variables (Browne & Nesselroade, 2005; Molenaar, 1985; Molenaar & Nesselroade, 2001; Zhang & Browne, 2010). Given that the optimal lag structures for the same data set may differ depending on whether lag relations are assumed among manifest variables or as exerted through latent factors, one important

extension to the present work would be to extend the proposed methods to facilitate the diagnosis of lag structures in models involving latent variables. Possible ways to accommodate latent variables may include the incorporation of a third stage in DTVEM to estimate latent variable scores that are then used within the exploratory GAMM stage, followed by the specification of state-space models in the confirmatory stage that can accommodate latent variables.

In summary, current exploratory methods to identify optimal times at which processes predict one another are inadequate for multisubject data and data with substantial missingness. In response to these inadequacies, we set out to design an exploratory model that would address these concerns. The principal utility of this tool is to provide a user-accessible manner to uncover and estimate optimal lag orders in a state-space framework. On the basis of this model's performance in simulation studies and the utility demonstrated in the empirical example, DTVEM appears to fulfill these needs.

Appendix: Example R code to run DTVEM

Load the DTVEM Package

```
library(DTVEM)
```

Load the data

```
data(example1dat)
```

Run the model

```
out = LAG("X", differentialtimevaryingpredictors = c("X"),  
outcome = c("X"), data = example1dat1, Time = "Time", k=9,  
ID="ID")
```

Summarize the results

```
summary(out)
```

See the options in the DTVEM Function

```
?LAG
```

Default Model Fitted

```
LAG(..., differentialtimevaryingpredictors = NULL, outcome =  
NULL,  
controlvariables = NULL, data = NULL, ID = "ID", Time  
= NULL,  
k = 10, k2 = 10, k3 = 3, k4 = 3, controllag = NULL,  
standardized = TRUE, predictionstart = NULL,  
predictionend = NULL,  
predictionsinterval = NULL, software = "OpenMx",  
independentpredictors = FALSE,  
minimumpracticalsignificance = NULL,  
gamma = 1, minN = 30, debug = FALSE,  
OpenMxStartingValues = 0.3,  
ResidualAnalysis = "Group", blockdata = FALSE,  
rounddecimals = TRUE,  
maxintermediaterounds = 10)
```

Descriptions of Primary Options in DTVEM

differentialtimevaryingpredictors

The variables that will be a varying-coefficient of differential time, meaning that this defines the variables used as $x_{1,lag,i,l}^*$ in $f_{2,l,1}(\Delta t_{1i,l}^*)x_{1,lag,i,l}^*$. These are the variables that one is interested in ascertaining the optimal lag structures in which a construct predicts the outcome(s). This must be specified as a vector using *c("variables here")*. e.g. *c("X", "Y")*. (REQUIRED).

outcome

This is each of the outcome variables. Specified as *outcome="outcomevariablename"* for a single variable or *outcome=c("outcomevariablename1", "outcomevariablename2")* (REQUIRED)

data

Specify the data frame that contains the data e.g. *data=dataframename* (REQUIRED)

ID

The name of the ID variable. E.G. *ID = "ID"* (must be specified). (REQUIRED)

Time

The name of the Time variable. E.G. *Time = "Time"* (must be specified). (REQUIRED)

k, k3, k4

The term *k* refers to the number of *k* selection points used in the model for the exploratory stage, as informed by diagnostic statistics from mgcv (see ?choose.k in mgcv package for more details; the ideal *k* may be selected on the basis of evaluations of diagnostic statistics from mgcv). The term *k3* refers to the *k* selection points used in the model for the nonparametric time trend (i.e., controlling for the time trend of the population). The term *k4* refers to the number of *k* selection points used in the model for the varying coefficient in the later iterations of the exploratory stage. It is safe to include the smallest *k* for which the “kindex” output by the DTVEM function is no longer significant. (OPTIONAL, BUT RECOMMENDED).

standardized

This specifies whether all of the variables (aside from Time) should be standardized. Options are *TRUE*, *FALSE*, and *"center"*. *TRUE* means within-person standardize each variable (a.k.a/ get the person-centered z-scores), *FALSE* means use the raw data. "Center" means to only within-person mean-center the variables. Default = *TRUE*. *FALSE* is not recommended unless you have done these transformations yourself (OPTIONAL)

ResidualAnalysis

Only applies when software = "OpenMx". Analyze the residuals of the time series with OpenMx after factoring out the non-linear effect of time (takes time trends into account). Can be run only at the group level (faster), or it can also be run with a random effect splines of time (slower) by setting *ResidualAnalysis = "Individual"*. Default = *"Group"* (OPTIONAL)

References

- Ahmad, I., Leelahanon, S., & Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Annals of Statistics*, 33, 258–283. <https://doi.org/10.1214/009053604000000931>
- Armstrong, A. W., Ulukan, S., Weiner, M., Mostafa, M., Shaheen, H., Nakhla, I., . . . Riddle, M. S. (2010). A randomized, double-blind, placebo-controlled study evaluating the efficacy and safety of rifaximin for the prevention of travelers' diarrhea in US military personnel deployed to Incirlik Air Base, Incirlik, Turkey. *Journal of Travel Medicine*, 17, 392–394. <https://doi.org/10.1111/j.1708-8305.2010.00462.x>
- Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). Estimation with applications to tracking and navigation: Theory algorithms and software. New York, NY: Wiley.
- Barrett, L. F., Quigley, K. S., Bliss-Moreau, E., & Aronson, K. R. (2004). Interoceptive sensitivity and self-reports of emotional experience. *Journal of Personality and Social Psychology*, 87, 684–697. <https://doi.org/10.1037/0022-3514.87.5.684>
- Beidel, D. C., Neal, A. M., & Lederer, A. S. (1991). The feasibility and validity of a daily diary for the assessment of anxiety in children. *Behavior Therapy*, 22, 505–517. [https://doi.org/10.1016/S0005-7894\(05\)80342-9](https://doi.org/10.1016/S0005-7894(05)80342-9)
- Boker, S. M., & Graham, J. (1998). A dynamical systems analysis of adolescent substance abuse. *Multivariate Behavioral Research*, 33, 479–507. https://doi.org/10.1207/s15327906mbr3304_3
- Bolger, N., & Laurenceau, J.-P. (2013). Intensive longitudinal methods: An introduction to diary and experience sampling research. New York, NY: Guilford Press.
- Bolkhovskiy, J. B., Scully, C. G., & Chon, K. H. (2012). *Statistical analysis of heart rate and heart rate variability monitoring through the use of smart phone cameras*. Paper presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), San Diego, CA.
- Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 567–585. <https://doi.org/10.1109/34.24792>

- Bottan, N. L., & Perez Truglia, R. (2011). Deconstructing the hedonic treadmill: Is happiness autoregressive? *Journal of Socio-Economics*, 40, 224–236. <https://doi.org/10.1016/j.socsec.2011.01.007>
- Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis, forecasting and control (Revised ed.). Englewood Cliffs, NJ: Prentice Hall.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2013). Time series analysis: Forecasting and control (4th ed.). Hoboken, NJ: Wiley.
- Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22, 409–425. <https://doi.org/10.1037/met0000085>
- Brockwell, P. J., & Davis, R. A. (2002). Introduction to time series and forecasting (2nd ed.). New York, NY: Springer.
- Brosschot, J. F., & Thayer, J. F. (2003). Heart rate response is longer after negative emotions than after positive emotions. *International Journal of Psychophysiology*, 50, 181–187.
- Brosschot, J. F., Van Dijk, E., & Thayer, J. F. (2007). Daily worry is related to low heart rate variability during waking and the subsequent nocturnal sleep period. *International Journal of Psychophysiology*, 63, 39–47. <https://doi.org/10.1016/j.ijpsycho.2006.07.016>
- Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. In O. Maydeu & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 415–452). Mahwah, NJ: Erlbaum.
- Buckner, J. D., Crosby, R. D., Wonderlich, S. A., & Schmidt, N. B. (2012). Social anxiety and cannabis use: An analysis from ecological momentary assessment. *Journal of Anxiety Disorders*, 26, 297–304.
- Buyssse, D. J., Thompson, W., Scott, J., Franzen, P. L., Germain, A., Hall, M., . . . Kupfer, D. J. (2007). Daytime symptoms in primary insomnia: A prospective analysis using ecological momentary assessment. *Sleep Medicine*, 8, 198–208.
- Cannon, W. B. (1927). The James–Lange theory of emotions: a critical examination and an alternative theory. *American Journal of Psychology*, 39, 106–124. <https://doi.org/10.2307/1415404>
- Carels, R. A., Musher-Eizenman, D., Cacciapaglia, H., Perez-Benitez, C. I., Christie, S., & O'Brien, W. (2004). Psychosocial functioning and physical symptoms in heart failure patients: A within-individual approach. *Journal of Psychosomatic Research*, 56, 95–101. [https://doi.org/10.1016/S0022-3999\(03\)00041-2](https://doi.org/10.1016/S0022-3999(03)00041-2)
- Chatfield, C. (2013). *The analysis of time series: An introduction*: CRC press.
- Chow, S.-M., & Zhang, G. (2013). Nonlinear regime-switching state-space (RSSS) models. *Psychometrika*, 78, 740–768. <https://doi.org/10.1007/s11336-013-9330-8>
- Chow, S.-M., Hamagani, F., & Nesselroade, J. R. (2007). Age differences in dynamical emotion–cognition linkages. *Psychology and Aging*, 22, 765–780. <https://doi.org/10.1037/0882-7974.22.4.765>
- Chow, S.-M., Haltigan, J. D., & Messinger, D. S. (2010). Dynamic patterns of infant–parent interactions during Face-to-Face and Still-Face episodes. *Emotion*, 10, 101–114. <https://doi.org/10.1037/a0017824>
- Chow, S.-M., Ho, M.-H. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalences and differences between structural equation and state-space modeling frameworks. *Structural Equation Modeling*, 17, 303–332. <https://doi.org/10.1080/10705511003661553>
- Chow, S.-M., Zu, J., Shifren, K., & Zhang, G. (2011). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Research*, 46, 303–339. <https://doi.org/10.1080/00273171.2011.563697>
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). New York, NY: Routledge.
- Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., & Goodner, B. (2007). Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics*, 16, 265–288. <https://doi.org/10.1198/106186007X208768>
- Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, 11, 1–8. <https://doi.org/10.1186/1741-7015-11-126>
- de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- versus continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, 8, 1849. <https://doi.org/10.3389/fpsyg.2017.01849>
- Durbin, J., & Koopman, S. J. (2012). Time series analysis by state space methods. Oxford, UK: Oxford University Press.
- Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology*, 118, 195–202. <https://doi.org/10.1037/a0014868>
- Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. J. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science*, 6, 416–427. <https://doi.org/10.1177/2167702617744325>
- Frank, B., Jacobson, N. C., Hurley, L., & McKay, D. (2017). A theoretical and empirical modeling of anxiety integrated with RDoC and temporal dynamics. *Journal of Anxiety Disorders*, 51, 39–46. <https://doi.org/10.1016/j.janxdis.2017.09.002>
- Gasparrini, A., Armstrong, B., & Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29, 2224–2234. <https://doi.org/10.1002/sim.3940>
- Gates, K. M., Molenaar, P. C., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage*, 50, 1118–1125. <https://doi.org/10.1016/j.neuroimage.2009.12.117>
- Goldstein, H., Healy, M. J. R., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, 13, 1643–1655. <https://doi.org/10.1002/sim.4780131605>
- Guttman, L., & Guttman, R. (1965). A new approach to the analysis of growth patterns: The simplex structure of intercorrelations of measurements. *Growth*, 29, 219–232.
- Hamilton, J. D. (1994). Time series analysis. Princeton, NJ: Princeton University Press.
- Harvey, A. C. (2001). Forecasting, structural time series models and the Kalman filter. Cambridge, UK: Cambridge University Press.
- Hastie, T. J., & Tibshirani, R. J. (1990). Generalized additive models. London, UK: Chapman & Hall.
- Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B*, 55, 757–796. <https://doi.org/10.2307/2345993>
- Ho, M.-H. R., Shumway, R., & Ombao, H. (2006). The state-space approach to modeling dynamic processes. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data*. (pp. 148–175). New York, NY: Oxford University Press.
- Hodges, W. F. (2015). The psychophysiology of anxiety. In M. Zuckerman & C. D. Spielberger (Eds.), *Emotions and anxiety: New concepts, methods, and applications* (pp. 175–194). New York, NY: Psychology Press.
- James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B*, 64, 411–432. <https://doi.org/10.1111/1467-9868.00342>

- Kelley, K. (2007). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods*, 39, 979–984. <https://doi.org/10.3758/bf03192993>
- Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling*. Advance online publication. <https://doi.org/10.1080/10705511.2018.1431046>
- Lang, P. J., McTeague, L. M., & Bradley, M. M. (2016). RDoC, DSM, and the reflex physiology of fear: A bio-dimensional analysis of the anxiety disorders spectrum. *Psychophysiology*, 53, 336–347. <https://doi.org/10.1111/psyp.12462>
- Lange, C. G., & James, W. (1922). *The emotions* (Vol. 1). Baltimore, MD: Williams & Wilkins.
- MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, 32, 215–253. https://doi.org/10.1207/s15327906mbr3203_1
- Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., . . . Maris, G. (2018). An introduction to network psychometrics: Relating using network models to item response theory models. *Multivariate Behavioral Research*, 53, 15–35. <https://doi.org/10.1080/00273171.2017.1379379>
- Mauss, I., Wilhelm, F., & Gross, J. (2004). Is there less to social anxiety than meets the eye? Emotion experience, expression, and bodily responding. *Cognition and Emotion*, 18, 631–642. <https://doi.org/10.1080/02699930341000112>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London, UK: Chapman & Hall.
- McKeown, G. J., & Sneddon, I. (2014). Modeling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychological Methods*, 19, 155–174. <https://doi.org/10.1037/a0034282>
- Mergler, D., & Valciukas, J. (1998). Nervous system: Overview. In J. M. Stellman (Ed.), *Encyclopaedia of occupational health and safety* (4th ed.). Geneva, Switzerland: International Labour Office.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., . . . Orwant, J. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331, 176–182.
- Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50, 181–202. <https://doi.org/10.1007/BF02294246>
- Molenaar, P. C. M., & Nesselroade, J. R. (2001). Rotation in the dynamic factor modeling of multivariate stationary time series. *Psychometrika*, 66, 99–107. <https://doi.org/10.1007/bf02295735>
- Morris, L. W., & Liebert, R. M. (1970). Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance. *Journal of Consulting and Clinical Psychology*, 35, 332–337. <https://doi.org/10.1037/h0030132>
- Muirhead, R. J. (1985). Estimating a particular function of the multiple correlation coefficient. *Journal of the American Statistical Association*, 80, 923–925. <https://doi.org/10.1080/01621459.1985.10478204>
- 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*, 81, 535–549. <https://doi.org/10.1007/s11336-014-9435-8>
- Nesselroade, J. R., & Ford, D. H. (1985). P-technique comes of age. Multivariate, replicated, single-subject designs for research on older adults. *Research on Aging*, 7, 46–80.
- Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. (2001). Dynamic factor analysis models for representing process in multivariate time-series. In D. M. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 235–265). Mahwah, NJ: Erlbaum.
- Oud, J. (2004). SEM state space modeling of panel data in discrete and continuous time and its relationship to traditional state space modeling. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models* (Vol. 19, pp. 13–40). Amsterdam, The Netherlands: Springer. https://doi.org/10.1007/978-1-4020-1958-6_2
- Reinsel, G. C. (2003). *Elements of multivariate time series analysis* (Springer Series in Statistics). New York, NY: Springer.
- Roche, M. J., Jacobson, N. C., & Pincus, A. L. (2016). Using repeated daily assessments to uncover oscillating patterns and temporally-dynamic triggers in structures of psychopathology: Applications to the DSM-5 alternative model of personality disorders. *Journal of Abnormal Psychology*, 125, 1090–1102. <https://doi.org/10.1037/abn0000177>
- Scully, C., Jinseok, L., Meyer, J., Gorbach, A. M., Granquist-Fraser, D., Mendelson, Y., & Chon, K. H. (2012). Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering*, 59, 303–306. <https://doi.org/10.1109/TBME.2011.2163157>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Shiyko, M. P., Lanza, S. T., Tan, X., Li, R., & Shiffman, S. (2012). Using the time-varying effect model (TVEM) to examine dynamic associations between negative affect and self confidence on smoking urges: Differences between successful quitters and relapsers. *Prevention Science*, 13, 288–299. <https://doi.org/10.1007/s11212-011-0264-z>
- Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned missing data designs in experience sampling research: Monte Carlo simulations of efficient designs for assessing within-person constructs. *Behavior Research Methods*, 46, 41–54. <https://doi.org/10.3758/s13428-013-0353-y>
- Starr, L. R., & Davila, J. (2012). Temporal patterns of anxious and depressed mood in generalized anxiety disorder: A daily diary study. *Behaviour Research and Therapy*, 50, 131–141.
- Steger, M. F., & Frazier, P. (2005). Meaning in life: One link in the chain from religiousness to well-being. *Journal of Counseling Psychology*, 52, 574–582. <https://doi.org/10.1037/0022-0167.52.4.574>
- Turchin, P., & Taylor, A. D. (1992). Complex dynamics in ecological time series. *Ecology*, 73, 289–305. <https://doi.org/10.2307/1938740>
- 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*, 66, 103–126. <https://doi.org/10.1111/j.2044-8317.2012.02043.x>
- Voelkle, M. C., Oud, J. H., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, 17, 176–192. <https://doi.org/10.1037/a0027543>
- Walls, T. A., & Schafer, J. L. (2006). *Models for intensive longitudinal data*. New York, NY: Oxford University Press.
- Wei, W. W. S. (2006). *Time series analysis: Univariate and multivariate methods* (2nd ed.). Boston, MA: Pearson Education, Inc.
- Wetherell, J. (2013). *Android heart rate monitor*. Retrieved from <https://code.google.com/p/android-heart-rate-monitor/>
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B*, 65, 95–114. <https://doi.org/10.1111/1467-9868.00374>
- Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: CRC Press.
- Zhang, G., & Browne, M. (2010). Dynamic factor analysis with ordinal variables. In S.-M. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (Notre Dame Series on Quantitative Methodology), Vol. 4, pp. 241–263. New York, NY: Taylor & Francis.