



Dartmouth
GEISEL SCHOOL OF
MEDICINE

DEPARTMENT OF BIOMEDICAL
DATA SCIENCE
GEISEL SCHOOL OF MEDICINE



Center for **Technology**
and **Behavioral Health**
Innovate · Evaluate · Disseminate

The Differential Time-Varying Effect Model (DTVEM):

A Tool for Diagnosing Optimal Measurement and Modeling Intervals in Longitudinal Data

Nicholas C. Jacobson, Ph.D.

Summer Data Science
and AI Webinar Series
July 2020



@NC_Jacobson

nicholasjacobson.com

Popularity of Longitudinal Assessment

- Intensive longitudinal data=densely collected data with 5+ measurements of the same variable
 - Typically 10-300 in behavioral active data collection within behavioral sciences
- Collection and analysis of intensive longitudinal data has grown exponentially
 - 385% growth in “experience sampling assessment” from 1990 to 2008 Google Ngram Viewer (online phrase usage tool)
- Often irregularly spaced time intervals for behavioral data
 - I.E. During the night, design reasons, following the occurrence of an event

Popular Methods of Analyzing Longitudinal Data

- Examples of Models Used in Analyzing Longitudinal Data:
 - Linear and non-linear time series models (including variations of vector autoregressive moving average (VARMA) models)
 - state-space models
 - dynamical systems models
 - multilevel modeling
 - and many other examples.¹
- Useful when researchers have **preconceived** notions about the number of lagged occasions to include in models

1. (Boker & Graham, 1998; Box, Jenkins, & Reinsel, 2013; Browne & Nesselroade, 2005; Durbin & Koopman, 2012; Goldstein, Healy, & Rasbash, 1994)

Choosing Lags due to convenience

- The vast majority of intensive longitudinal data studies that utilize dynamic relationships only look at lower-order lags
 - Typically lags of 1
- This strategy assumes that higher order lags are negligible
 - This **NOT** an innocuous assumption
- When higher order lags exist, failing to account for higher-order lags leads to:
 - Misleading inferential results (i.e. biased parameter estimates, with underestimated standard errors)

Choose Time Lags

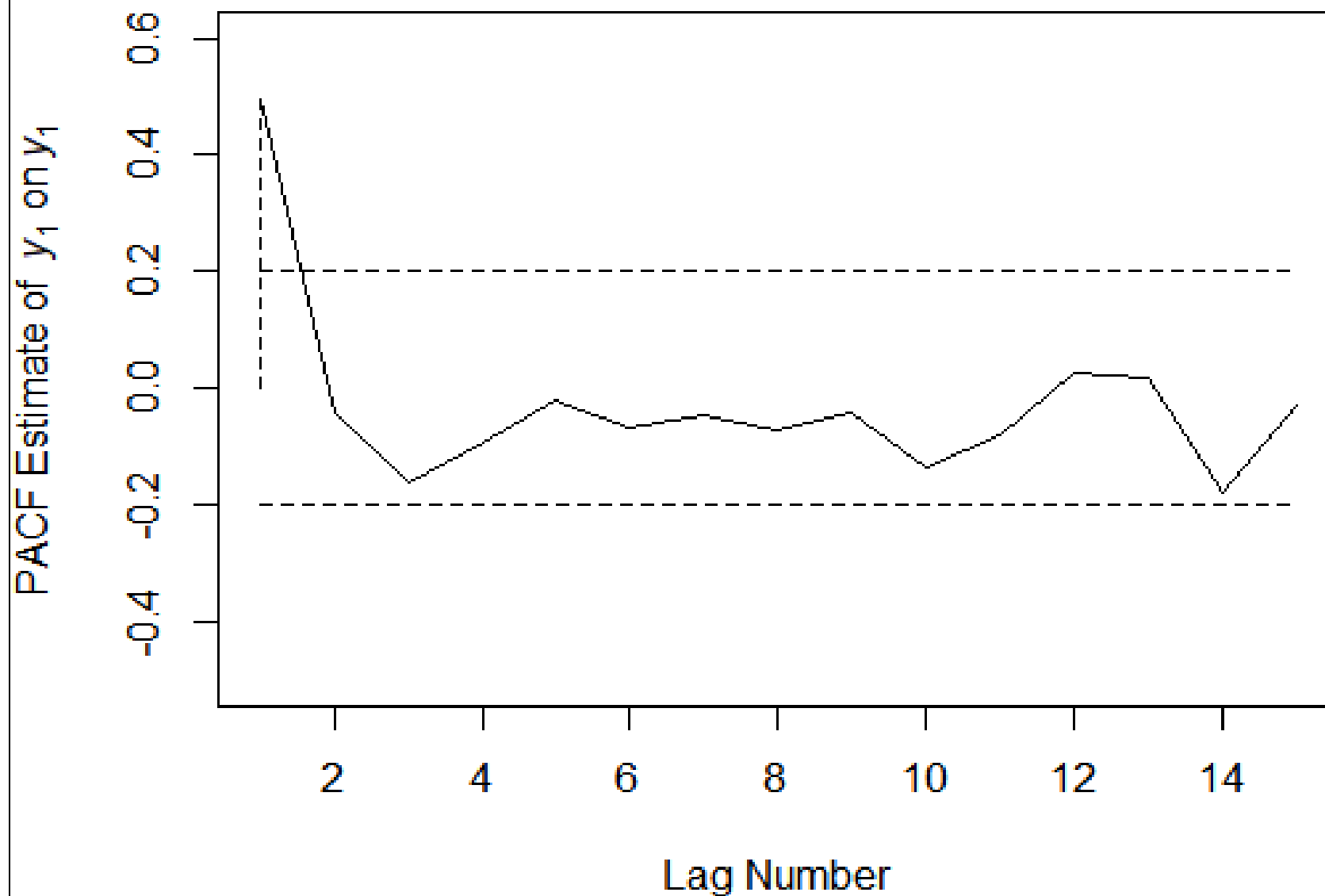
- Methods to choose lag times
 - Traditionally chosen in an exploratory fashion using diagnostics such as auto- and partial correlation plots
 - In a confirmatory way as guided by theories
 - Comparing results from model fitting at different lags¹

1. (e.g., Chow, Nesselroade, Shifren, & McArdle, 2004)

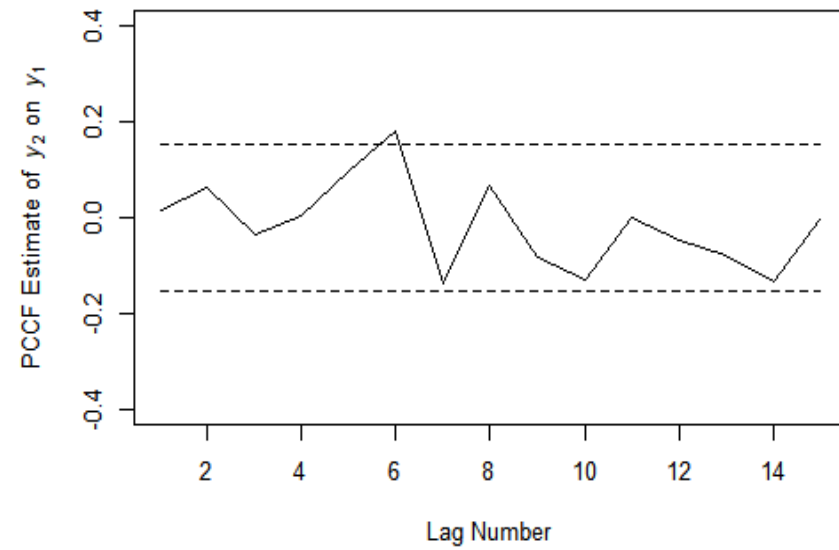
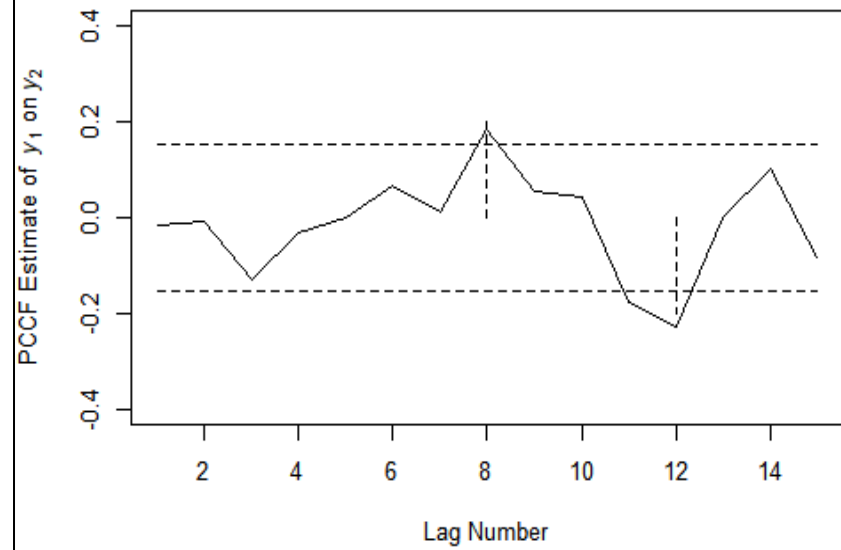
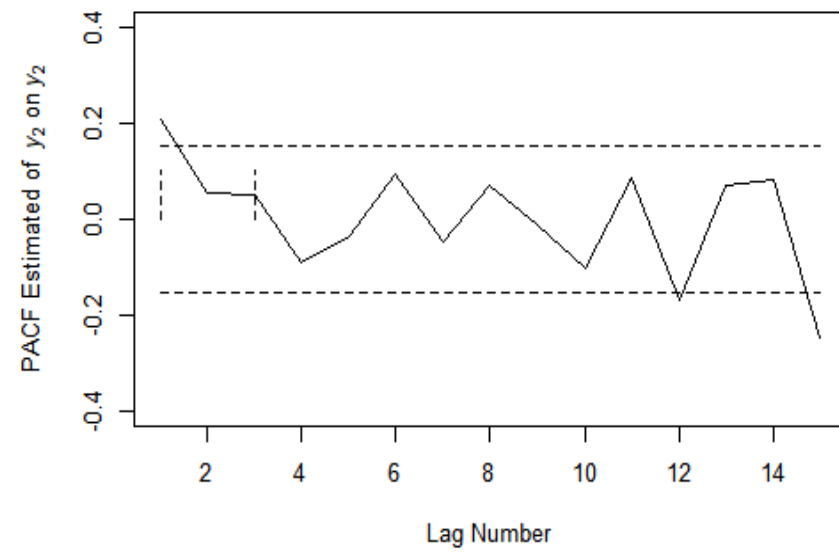
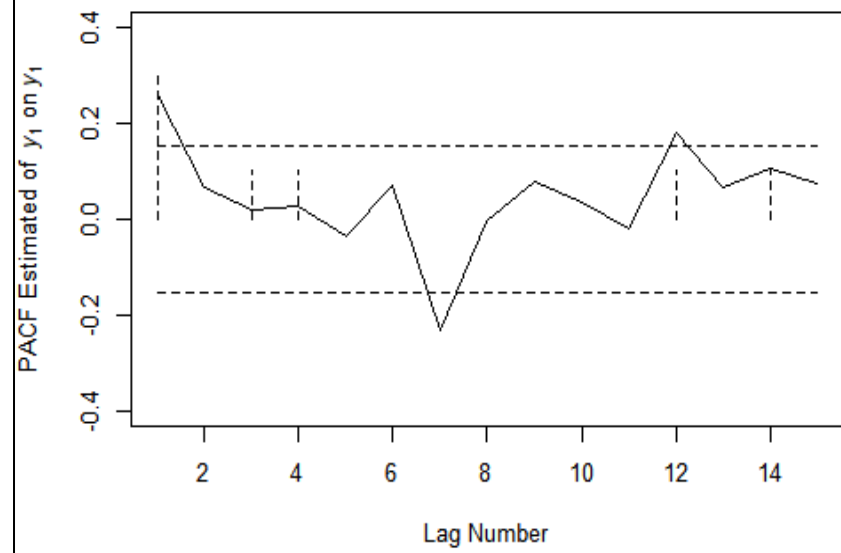
Inadequacies of Previous Methods for Behavioral Science Data

1. Traditional exploratory methods have been designed for:
 - Single-subject time series data of more substantial lengths ($t \geq 100$)
 - (e.g. with more than 100 time points and no missingness)
2. Confirmatory methods:
 - Computationally inefficient with large amount of lags

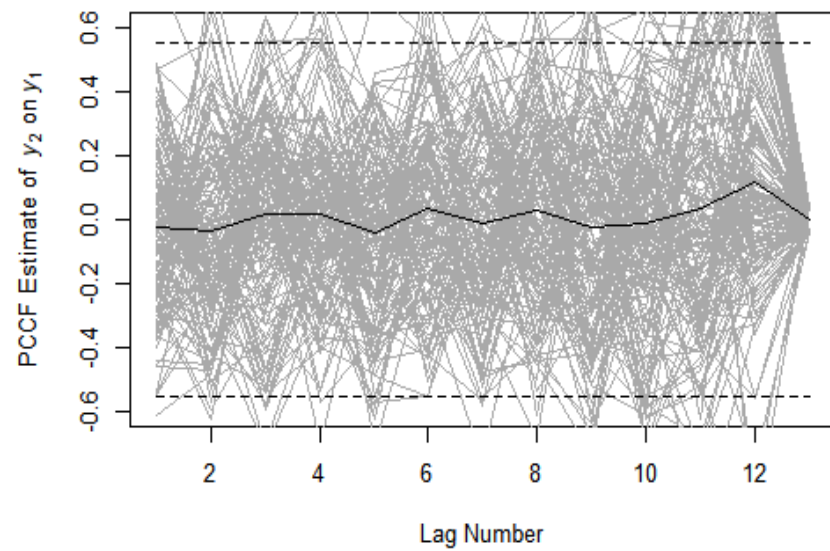
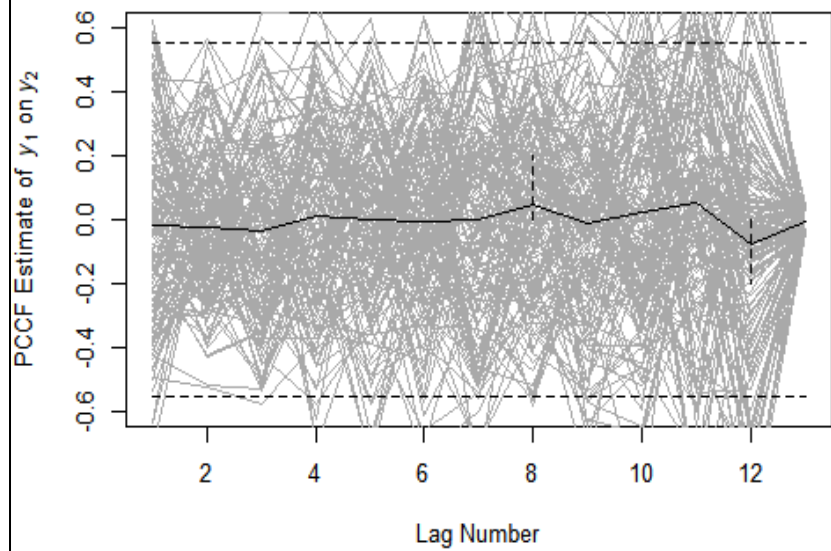
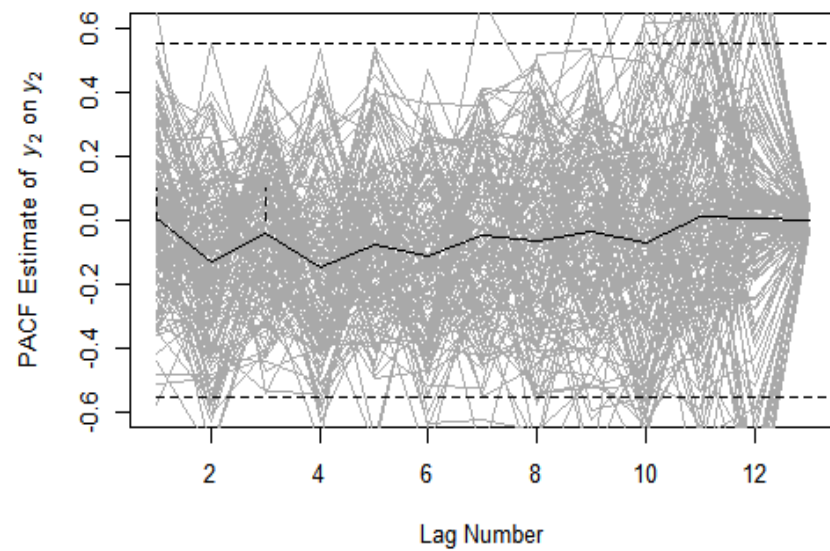
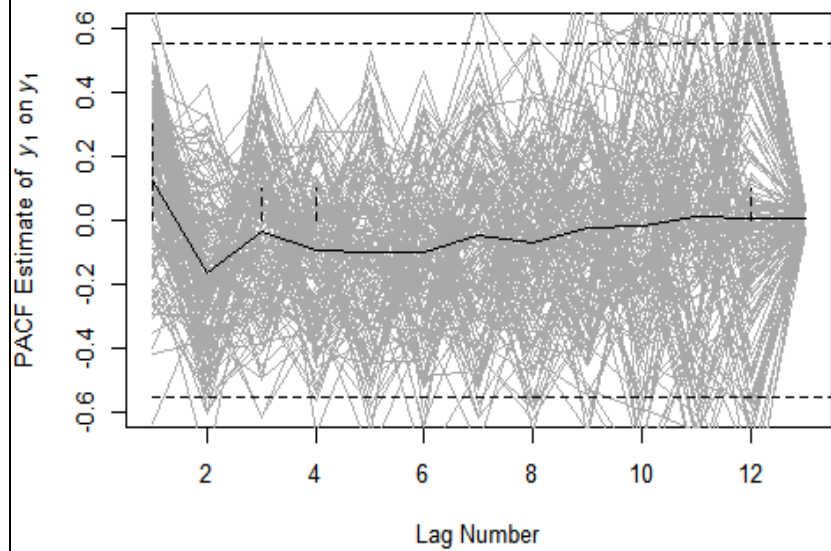
1a: Illustration 1 $N = 1$ $T = 100$, No Missing Data



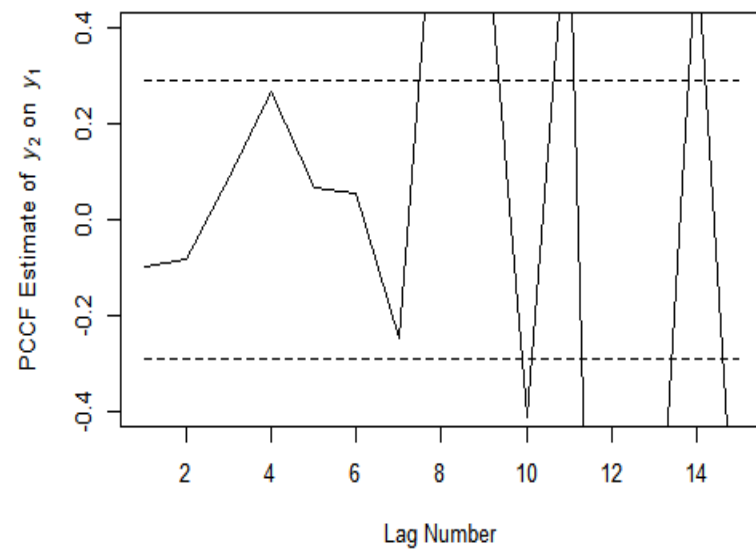
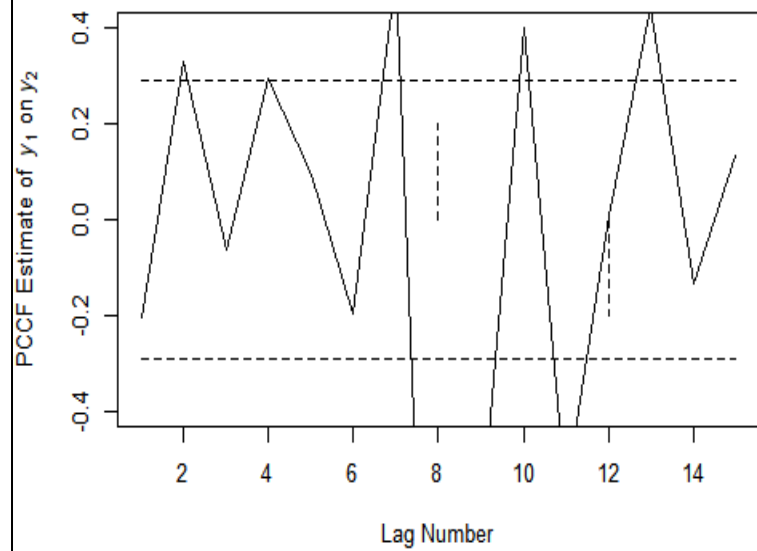
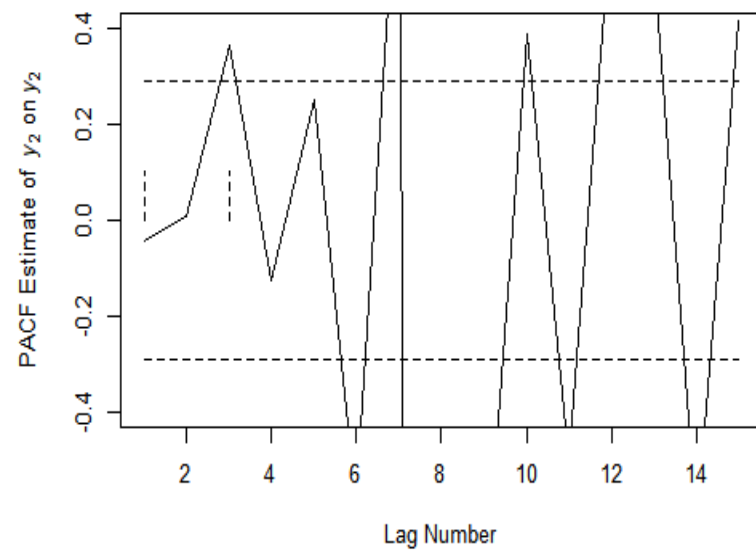
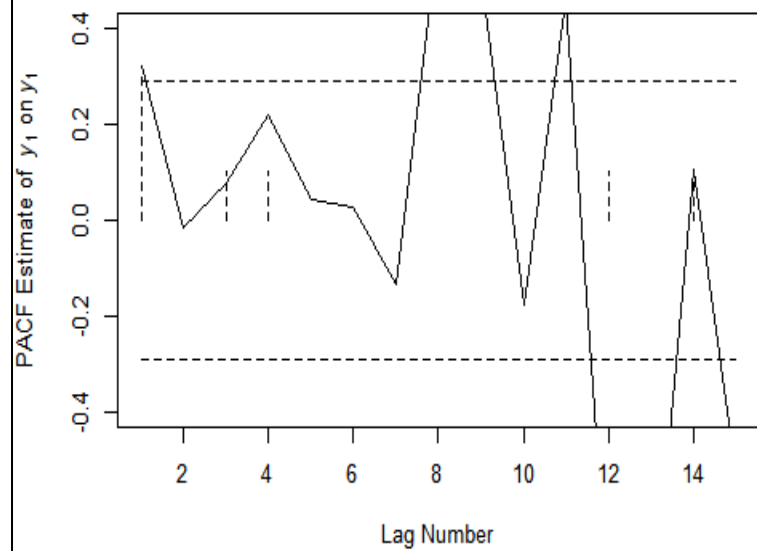
1b: Illustration 2 $N=1$ $T=170$, No Missing Data



1c: Illustration 3 $N=159$ $T=14$, No Missing Data



1d: Illustration 4 $N=1$ $T=170$, 72% Missing Data



Current Approach

- The current presentation:
 - Novel exploratory approach for identifying optimal lags in multiple-subject, multivariate (or univariate) longitudinal data
- This new approach that extends previous approaches by:
 - Allowing for the fitting of automated exploratory and confirmatory modeling of higher-order lags through a two-stage model fitting procedure.

Differential Time Varying Effect Model

- DTVEM is a tool for diagnosing the optimal time lag intervals in equally and unequally spaced ILD
- Capitalizes on some of the flexible and smoothing estimation routines for generalized additive mixed model (GAMM)
- Followed up with the strong estimation confirmatory modeling performance with vector autoregressive models within a state-space routine


Exploratory Stage Model Formula

$$\mu_i \equiv E(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki}) x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$

Exploratory Stage Model Formula

$$\begin{aligned} \mu_i &\equiv \mathbf{E}(y_i) \\ \mu_i &= g^{-1}(\eta_i) \\ \eta_i &= \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i \end{aligned}$$


- person i 's response variable, y_i (where $i = 1, \dots, n$, with n indexing the total number of subjects)
- may be distributed as any of the members from the exponential family (e.g., Normal, Poisson, Gamma, Multinomial, etc.;

Exploratory Stage Model Formula


$$\begin{aligned} \mu_i &\equiv \mathbf{E}(y_i) \\ \mu_i &= g^{-1}(\eta_i) \\ \eta_i &= \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki}) x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i \end{aligned}$$

- Mean function

Exploratory Stage Model Formula

$$\mu_i \equiv \mathbf{E}(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

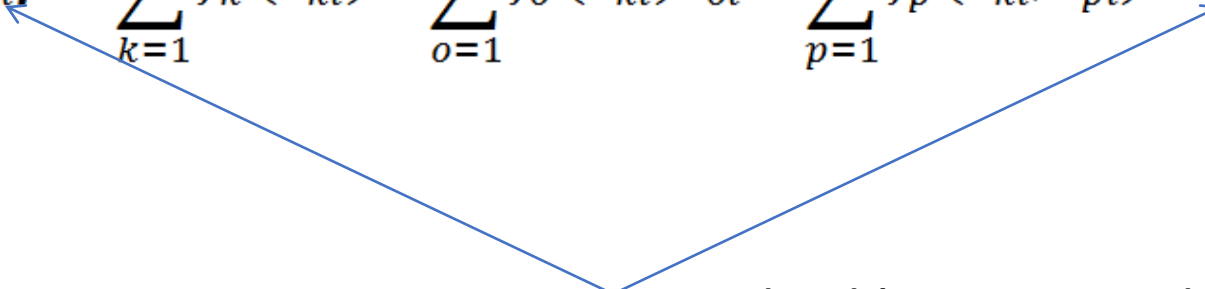
$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$


- link function that maps the mean of y_i into the semiparametric predictor and $g^{-1}(\eta_i)$ in the transformation in the reverse direction

Exploratory Stage Model Formula

$$\mu_i \equiv E(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

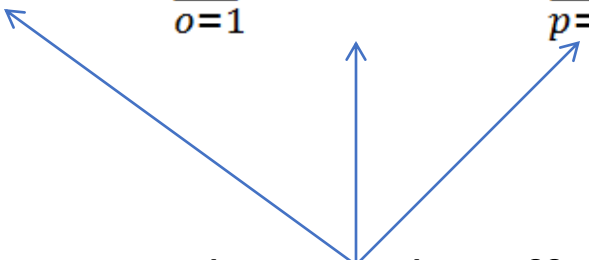
$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki}) x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$


- the usual parametric components in standard linear mixed effects models

Exploratory Stage Model Formula

$$\mu_i \equiv \mathbf{E}(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

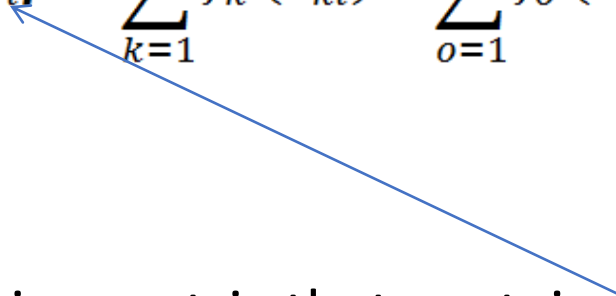
$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$


- nonparametric components wherein the effects of a series of covariates on the mean of the dependent variable are of unknown functional forms

Exploratory Stage Model Formula

$$\mu_i \equiv \mathbf{E}(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$


- \mathbf{X}_i is the design matrix that contains person i 's fixed effects components and $\boldsymbol{\beta}$ is the corresponding vector of fixed effects parameters

Exploratory Stage Model Formula

$$\boldsymbol{\mu}_i \equiv \mathbf{E}(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$


$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki}) x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$

- \mathbf{Z}_i is the random effects design matrix for person i and $\mathbf{b}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Psi}_b)$ assumed to be multivariate normally distributed with zero means and covariance matrix, $\boldsymbol{\Psi}_b$

Exploratory Stage Model Formula

$$\mu_i \equiv \mathbf{E}(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki}) x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$


- The term f_k is the smoothing function for the k th covariate, x_{ki} . ($k = 1, \dots, K$);

Exploratory Stage Model Formula

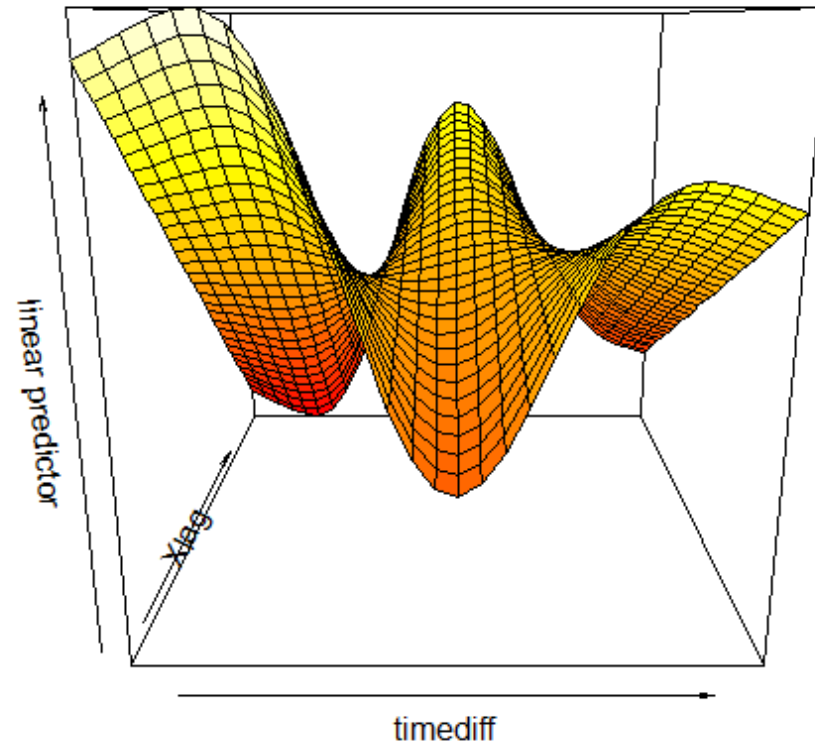
$$\mu_i \equiv \mathbf{E}(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki}) x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$

- f_o ($o = 1, \dots, O$) are O smoothing functions that allow the covariates in x_{oi} to have varying coefficients that may depend linearly or non-linearly on x_{ki}
 - x_{ki} may contain an indicator of time or space, in which case the effects of the covariates in x_{oi} on the mean function are assumed to be time- or space-varying¹

Varying Coefficient Example Plot



Exploratory Stage Model Formula

$$\boldsymbol{\mu}_i \equiv \mathbf{E}(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki}) x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$

- the terms f_p ($p = 1, \dots, P$) are P smooths of tensor products used to approximate the unknown effects of two interdependent covariates on η_i .
 - Means that both coefficients are allowed to vary non-linearly to predict eta

Exploratory Stage Model Formula

$$\mu_i \equiv \mathbf{E}(y_i)$$

$$\mu_i = g^{-1}(\eta_i)$$

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki}) x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$

- Link function in GAMM and therefore DTVEM by extension can handle non-normal data that conform to the exponential family
 - E.G. Count and categorical data

Setting Up the Data

ID	y_j	t_j
1	-0.403	1
1	0.200	2
1	-0.229	3
1	0.019	4
...
ID_i	y_{ij}	t_{ij}

Take the raw data and create lagged variables up to the total possible lag (l) for both the time (t) and the data of interest (y).

ID	y_j	y_{j-1}	y_{j-2}	y_{j-3}	...	y_{j-l}	t	t_{j-1}	t_{j-2}	t_{j-3}	...	t_{j-l}
1	-0.403	1
1	0.200	-0.403	2	1
1	-0.229	0.200	-0.403	3	2	1
1	0.019	-0.229	0.200	-0.403	4	3	2	1
...
ID_i	y_{ij}	y_{ij-1}	y_{ij-2}	y_{ij-3}	...	y_{ij-l}	t_{ij}	t_{ij-1}	t_{ij-2}	t_{ij-3}	...	t_{ij-l}

ID	y_j	y_{j-1}	Δt_{j-1}
1	-0.403	.	.
1	0.200	-0.403	1
1	-0.229	0.200	1
1	0.019	-0.229	1
1	-0.403	.	.
1	0.200	.	.
1	-0.229	-0.403	2
1	0.019	0.200	2
1	-0.403	.	.
1	0.200	.	.
1	-0.229	-0.403	3
1	0.019	-0.403	3
...
ID_i	y_{ij}	y_{ij-1}	Δt_{ij-1}

Take the lagged data matrix and stack it on top of one another for each lagged column (y_{j-1}), and subtract the lagged time variable (t_{j-1}) from the time (t) to create time differences $\Delta t_{i,j}$.

Exploratory Stage Model Formula

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$

- $y_i = [y_{i,j-1}, y_{i,j-2}, \dots, y_{i,j-J}]^T$ ←
 - $x_{ki} = [\Delta t_{i,j-1}, \Delta t_{i,j-2}, \dots, \Delta t_{i,j-J}]^T$
 - $x_{oi} = [x_{oi,j-1}, x_{oi,j-2}, \dots, x_{oi,j-J}]^T$
 - $x_{pi} = [x_{pi,j-1}, x_{pi,j-2}, \dots, x_{pi,j-J}]^T$
- y_i , is a vector consisting of all the stacked repeated measurements of the outcome variable up to, but not including, the measurement at time J

Exploratory Stage Model Formula

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$

- $y_i = [y_{i,j-1}, y_{i,j-2}, \dots, y_{i,j-J}]^T$
 - $x_{ki} = [\Delta t_{i,j-1}, \Delta t_{i,j-2}, \dots, \Delta t_{i,j-J}]^T$
 - $x_{oi} = [x_{oi,j-1}, x_{oi,j-2}, \dots, x_{oi,j-J}]^T$
 - $x_{pi} = [x_{pi,j-1}, x_{pi,j-2}, \dots, x_{pi,j-J}]^T$
- x_{ki} is the corresponding vector of time differences between the first and each of the measurements in y_i , and x_{oi} or x_{pi}

Exploratory Stage Model Formula

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$


- $y_i = [y_{i,j-1}, y_{i,j-2}, \dots, y_{i,j-J}]^T$
 - $x_{ki} = [\Delta t_{i,j-1}, \Delta t_{i,j-2}, \dots, \Delta t_{i,j-J}]^T$
 - $x_{oi} = [x_{oi,j-1}, x_{oi,j-2}, \dots, x_{oi,j-J}]^T$ ←
 - $x_{pi} = [x_{pi,j-1}, x_{pi,j-2}, \dots, x_{pi,j-J}]^T$
- x_{oi} is a vector of predictors of interest that will be used within a varying-coefficient framework as a function of time differences (x_{ki})

Exploratory Stage Model Formula

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki}, x_{pi}) + \mathbf{Z}_i \mathbf{b}_i$$

- $y_i = [y_{i,j-1}, y_{i,j-2}, \dots, y_{i,j-J}]^T$
- $x_{ki} = [\Delta t_{i,j-1}, \Delta t_{i,j-2}, \dots, \Delta t_{i,j-J}]^T$
- $x_{oi} = [x_{oi,j-1}, x_{oi,j-2}, \dots, x_{oi,j-J}]^T$
- $x_{pi} = [x_{pi,j-1}, x_{pi,j-2}, \dots, x_{pi,j-J}]^T$ ←
- x_{pi} is a vector of the predictor that will be used within the tensor-products with time differences

Exploratory Stage Model Formula

$$\eta_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{ki}) + \sum_{o=1}^O f_o(x_{ki})x_{oi} + \sum_{p=1}^P f_p(x_{ki})x_{pi} + \mathbf{Z}_i \mathbf{b}_i$$


- $y_i = [y_{i,j-1}, y_{i,j-2}, \dots, y_{i,j-J}]^T$
- $x_{ki} = [\Delta t_{i,j-1}, \Delta t_{i,j-2}, \dots, \Delta t_{i,j-J}]^T$
- $x_{oi} = [x_{oi,j-1}, x_{oi,j-2}, \dots, x_{oi,j-J}]^T$
- $x_{pi} = [x_{pi,j-1}, x_{pi,j-2}, \dots, x_{pi,j-J}]^T$
- In our experience, the varying-coefficient's estimates result in better estimates when the relationship between y_i and x_{oi} are a partially linear function of non-linear x_{ki} . Thus, our simulations have only utilized the second smooth term.

DTVEM Exploratory -> Confirmatory Stage

- Due to biased standard errors a more traditional confirmatory model is then undertaken
- To do this algorithm examines the local peaks and valleys
 - And all nearby values that are not significantly different from the local peaks and valleys
- All points grabbed need to be significantly different from zero to be passed on to traditional stage (vector autoregressive models)
- If there are no values that are identified, DTVEM notifies the user that there are no significant peaks and valleys within the time series.

Stage 2 - Detrending

$$\eta_i = \beta + \sum_{k=1}^K f_k(x_{ki}) + \epsilon_i$$

- $x_{ki} = [t_{i1}, t_{i2}, \dots, t_{ij}]^T$.

- x_{ki} is a vector of time
- Thus the effect of time on the response variable is allowed to vary non-linearly as a functional smooth

Confirmatory Model Stage

Confirmatory Vector autoregression (VAR(p)) model

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = A_1 \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + A_2 \begin{bmatrix} y_{1,i,t-2} \\ y_{2,i,t-2} \end{bmatrix} + A_3 \begin{bmatrix} y_{1,i,t-3} \\ y_{2,i,t-3} \end{bmatrix} + \varepsilon_{i,t}$$

Confirmatory Model Stage

Confirmatory Vector autoregression (VAR(p)) model

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = A_1 \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + A_2 \begin{bmatrix} y_{1,i,t-2} \\ y_{2,i,t-2} \end{bmatrix} + A_3 \begin{bmatrix} y_{1,i,t-3} \\ y_{2,i,t-3} \end{bmatrix} + \varepsilon_{i,t}$$

Outcomes

Lagged Regression weights

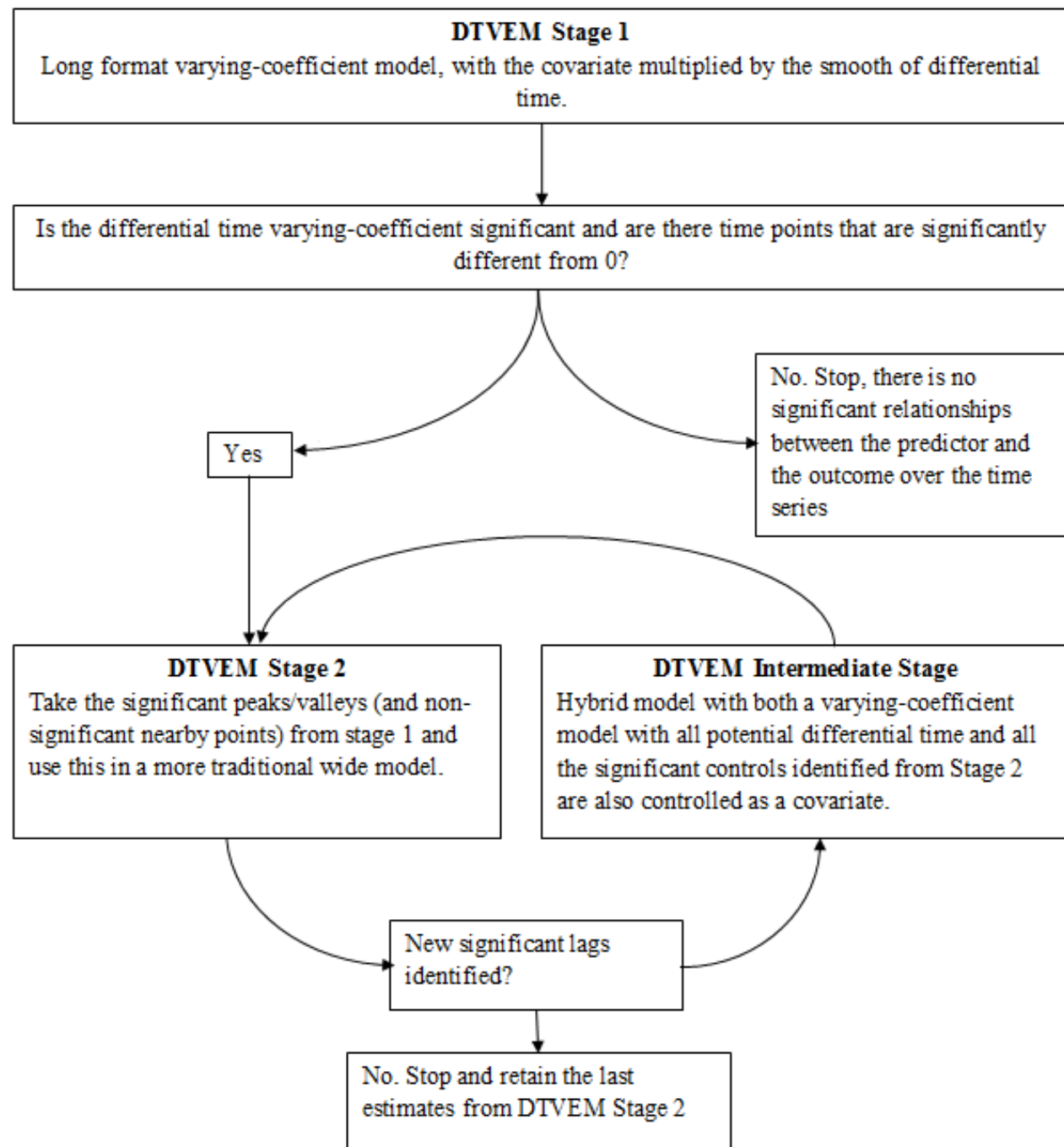


Further Exploratory Stage

- After Confirmatory:
- An intermediate stage is repeated to ensure that all of the potential lags are identified through a hybrid stage 1 and 2 model.
- In this intermediate stage, the matrices from stage 1 are identical.
- However, discrete lag times that were found to be significant in stage 2 are also controlled for through a fixed effects linear covariate

Following the Intermediate Stage

- If new peaks are identified, they are passed along to the previous confirmatory stage model but added as further covariates.
- If none of the peaks and valleys are significantly different from the rest of the time series or the peaks and valleys are not significantly different from zero, the stage 2 model estimates are retained.
- A loop is repeated until the intermediate stage is flat or it is unable to identify any new peaks and valleys.
- At the end of the DTVEM function, DTVEM returns the last stage 2 estimates.



Simulation Results

- Simulated data with a high percentage of missing data (up to 72%), low signal-to-noise ratios
 - Univariate and Multivariate data
 - Single subject and multi-subject time series
- DTVEM has high exploratory power and low Type I error rates across both univariate and bivariate models
 - Accurate point estimates and unbiased standard errors

Behavior Research Methods
<https://doi.org/10.3758/s13428-018-1101-0>



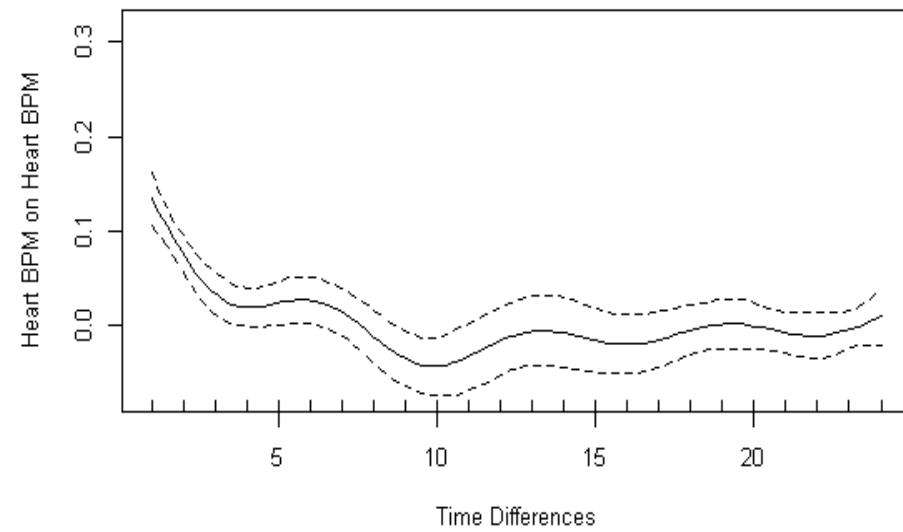
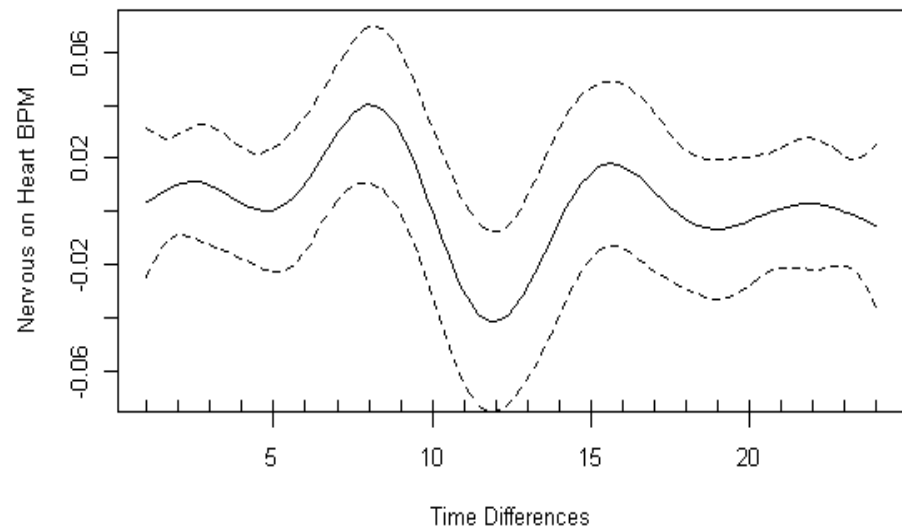
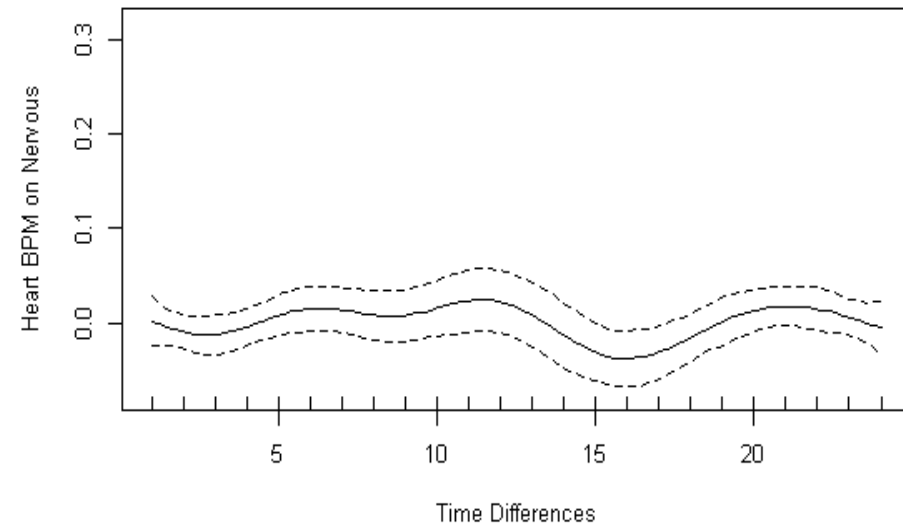
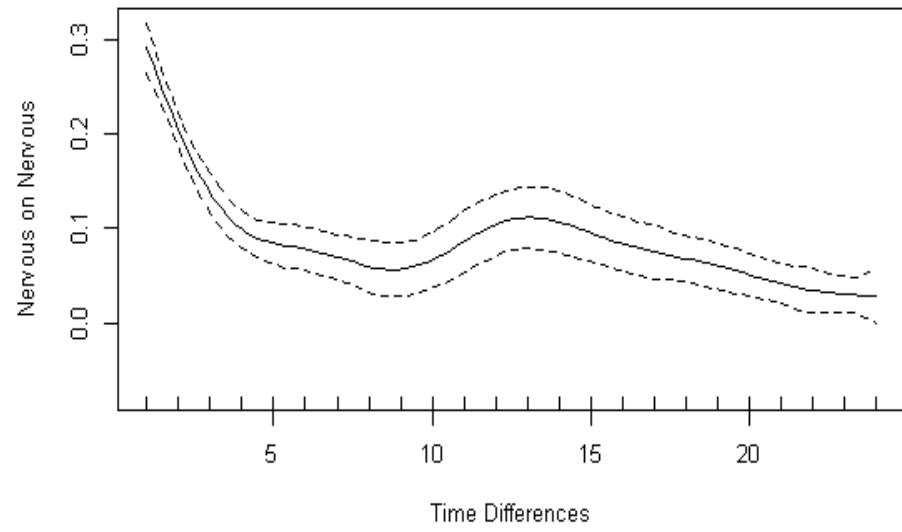
The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data

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

Empirical Example: Nervousness and Heart Rate

- EMA: Hourly assessments during waking times for one week
- Most participants ($N = 159$) completed 68.5% of prompts.
- The “nervous” item was used on a 0 – 100 slider.
 - Heart rate was measured with an open-source application which used the camera on smartphones (Wetherell, 2013).
- There was a total of 7,509 data points collected out of a possible 26,880 (if prompts had occurred evenly, and each person had a complete sampling of every period).
- Thus, the data were 28% complete.

DTVEM Stage 1



Confirmatory Estimates

- Nervousness predicted itself:
 - one hour later ($\alpha_{1,1} = 0.275, SE_{\alpha_{1,1}} = 0.012, p_{\alpha_{1,1}} < .001$)
 - three hours later ($\alpha_{1,3} = 0.071, SE_{\alpha_{1,3}} = 0.014, p_{\alpha_{1,3}} < .001$),
 - four hours later ($\alpha_{1,4} = 0.075, SE_{\alpha_{1,4}} = 0.014, p_{\alpha_{1,4}} < .001$),
 - 12 hours later ($\alpha_{1,12} = 0.077, SE_{\alpha_{1,12}} = 0.020, p_{\alpha_{1,12}} < .001$)
 - 14 hours later ($\alpha_{1,14} = 0.082, SE_{\alpha_{1,14}} = 0.019, p_{\alpha_{1,14}} < .001$)
- Heart rate significantly predicted itself:
 - one hour later ($\alpha_{2,1} = 0.144, SE_{\alpha_{2,1}} = 0.015, p_{\alpha_{2,1}} < .001$)
 - three hours later ($\alpha_{2,3} = 0.037, SE_{\alpha_{2,3}} = 0.016, p_{\alpha_{2,3}} = .020$)
- Nervousness significantly predicted heart rate:
 - Positively: 8 hours later ($\gamma_{1,2,8} = 0.055, SE_{\gamma_{1,2,8}} = 0.020, p_{\gamma_{1,2,8}} = .007$)
 - Negatively: 12 hours later ($\gamma_{1,2,12} = -0.052, SE_{\gamma_{1,2,12}} = 0.024, p_{\gamma_{1,2,12}} = .030$)

Empirical Example: Temporal Stability of Personality Impairments

- N = 248
- 14-day daily diary study
- Found evidence of oscillating personality impairment, and levels of personality dysfunction exacerbated the oscillating patterns

Journal of Abnormal Psychology
2016, Vol. 125, No. 8, 1090–1102

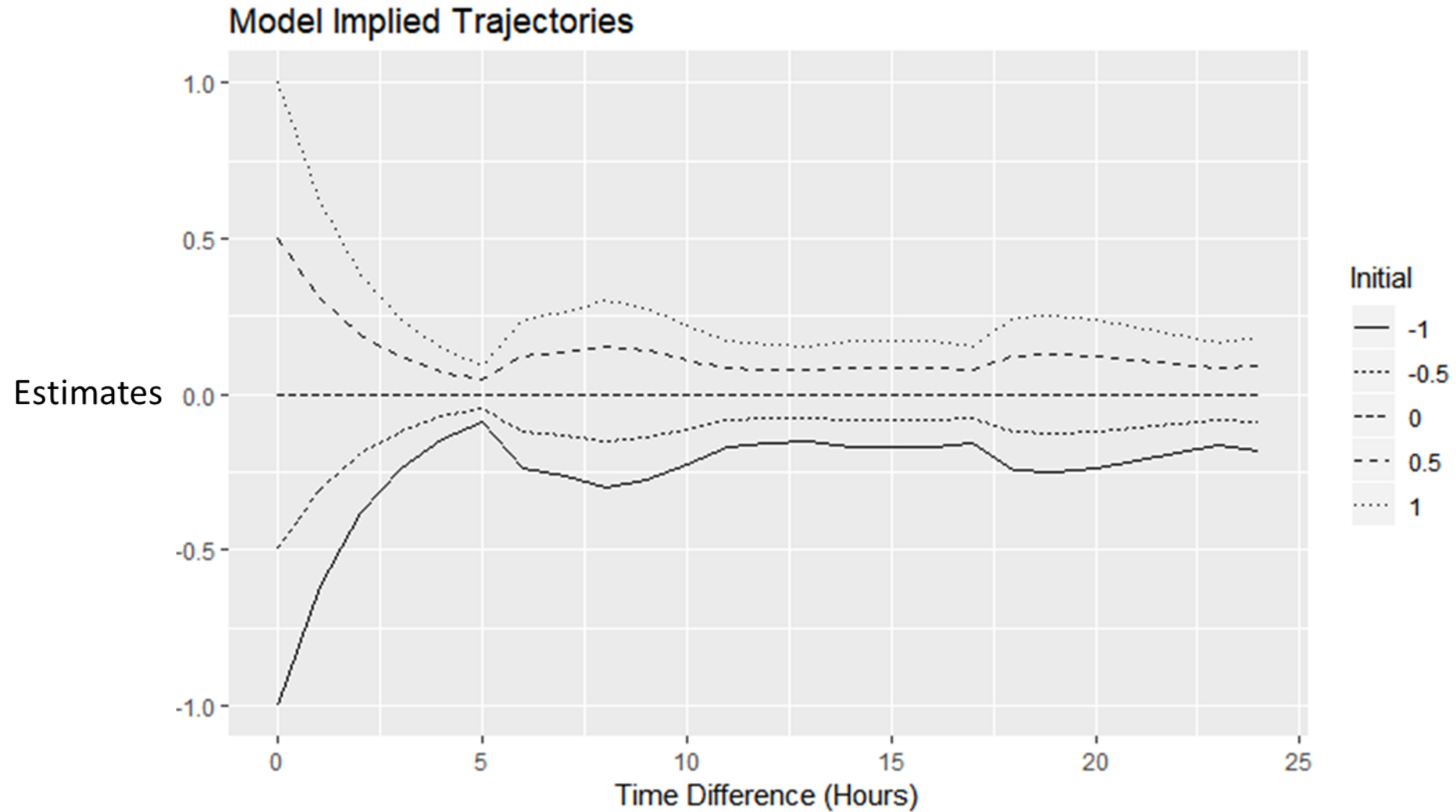
© 2016 American Psychological Association
0021-843X/16/\$12.00 <http://dx.doi.org/10.1037/abn0000177>

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

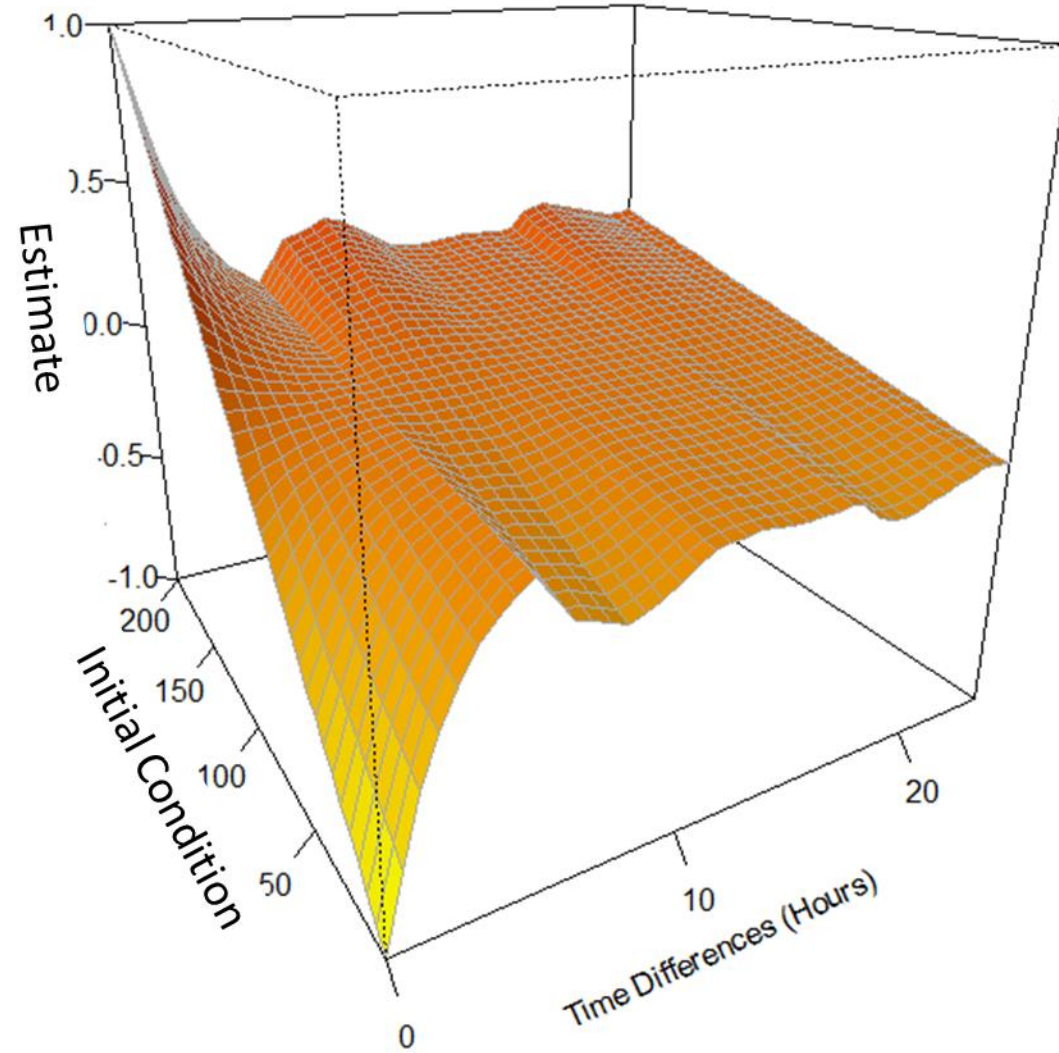
Michael J. Roche
Penn State Altoona

Nicholas C. Jacobson and Aaron L. Pincus
The Pennsylvania State University

Confirmatory Estimates



3D Plot of Confirmatory Estimates of Suicidalit



Summary

- DTVEM is a good method that is able to correctly identify true trajectories with high precision with small or large amounts of data points (as few as 14 per person)
 - Small ($N = 1$) or large sample sizes ($N > 1$)
 - And data with low signal to noise ratios (or higher)
- Thus, DTVEM is a promising and flexible approach that allows researchers to consider all the potential time relationships between a construct and itself or other constructs.

Installing DTVEM

- Install R
- Download the DTVEM R Package here:
 - <http://www.nicholasjacobson.com/project/dtvem/>
- Install the required dependencies:
 - `install.packages(c("mgcv","plyr","zoo","reshape2","Rcpp","OpenMx"))`
 - Note that in Windows you will also need to install Rtools
- Load DTVEM:
 - `library(DTVEM)`

Important inputs

- “X”
- differentialtimevaryingpredictors
- outcome
- controlvariables
- data
- ID
- Time
- predictionstart
- predictionsend
- Predictionsinterval

Useful inputs

- k, k_2, k_3, k_4
 - Controls the number of k selection points used in the model points – allows for more or less model bend in the exploratory stage
- standardized
 - Can person-standardize or person-center data for you
- ResidualAnalysis
 - Detrends the data for you
- gamma
 - Controls the penalization of wiggleness in the exploratory stage

Answering Emailed Questions

- What does k really mean?
 - k is a parameter that affects the number of degrees of freedom available for a smooth
 - Note that the upper number of k is $(t-1)$
 - This only affects the exploratory stage and not the confirmatory stage estimation
- If X predicts itself 8 time points later, does that mean in other words that the value at the 8th time point can be determined by the values at all time points leading up to the 8th?
 - This isn't quite right, if there is an autoregressive relationship at lag 8, then it suggests that a value predicts itself 8 time units later (it doesn't imply that there is also a relationship between a variable and itself 1-7 time units later if it only occurs at lag 8)

- How do the reported beta values correspond to the plots of beta coef vs time?
 - The plots that are depicted are from the exploratory stage of DTVEM
 - Note that these plots are of time differences (a continuous operationalization of a time “lag” which is discrete)
 - In a plot of X predicting X, this describes the relationship between X and itself across later time periods
 - This can give us a sense of the data, but the final time estimates should be used in drawing firm inferences
- What is an appropriate threshold for missingness?
 - This is entirely case dependent, and there’s no firm threshold for “too much” missingness. A longer time series can make up for greater missing values associated with the time series.
 - DTVEM can actually perform well in cases with ridiculous periods of missingness (i.e. 90% or more of data missing data) if there are enough time points otherwise

- For assessments two or three times per week, would it make more sense for the DTVEM input matrix to have weekly time points averaging the two values for a guaranteed value at each time point or daily time points where about half of the values are NA?
 - This would depend a bit on the a priori design and the desired estimation period.
 - Not average the values and just to record time in the unit of “days” though this would produce a large degree of missing data (one of the potential benefits would be greater precision in answering the smaller time-lagged relationships, but this could be problematic if it were systematically missing).
 - On the other hand, if there’s very little data over certain intervals (i.e. it’s always on Mondays and Thursdays), and not jittered, I would consider thresholding values in “half-weeks” as this will increase the precision of the smaller time intervals).
- I provide some functions to determine the effects of these blocking estimates in the LAG function, by having DTVEM do the blocking before and after. You can then look at the similarity/differences in the exploratory blocked estimates.