

Master's Paper of the Department of Statistics, the University of Chicago

Modeling Ecological Momentary Assessments Data Using Mixed-Effects Location-Scale Model and Time Varying Effects Model

Yan Zhang

Advisor: Donald Hedeker and Mei Wang

Approved: _____

Date: _____

Abstract

Ecological Momentary Assessment (EMA) studies make use of modern technologies such as smartphones to repeatedly sample subjects' behavior and state in real time. These studies produce intensively measured longitudinal data with large numbers of observations per subject, and are well suited to address research questions in psychology and public health while assuring a degree of ecological validity. This paper highlights two analytical approaches to EMA data. Mixed-effects location-scale jointly model the mean and variance of the response variable while accounting for individual heterogeneity. Time-varying effects model on the other hand estimates dynamic, nonlinear associations between variables by estimating coefficients as functions of time. Theoretical and analytical properties of these methods are described alongside simulation studies and real data analysis examples to illustrate their strengths and weaknesses in practice.

Contents

1	Introduction	3
2	Mixed-effects location-scale model	4
2.1	A motivating study	4
2.2	Formulation of MELS	4
2.3	Estimation of MELS model	5
2.3.1	Maximum likelihood estimation	6
2.3.2	Bayesian approaches	7
2.4	Simulation study of MELS	8
2.4.1	Impact of scale effect on inference	8
2.4.2	Robustness to misspecification of covariance structure	9
2.5	Mixed location-scale hidden Markov model	10
2.5.1	Motivation and model formulation	10
2.5.2	Estimation algorithm	11
2.5.3	Simulation study	12
2.6	Data analysis example	13
2.6.1	Null model for negative affect	13
2.6.2	Modeling of BS-variance	15
2.6.3	Full model for negative affect	15
2.6.4	Additional analysis	18
3	Time-varying effects models (TVEM)	19
3.1	Motivating study of TVEM	19
3.2	Formulation of TVEM	19
3.3	Estimation of TVEM	20
3.3.1	Truncated power basis	20
3.3.2	Penalizing model complexity	21
3.3.3	Dealing with dependence structure in data	22
3.3.4	Generalized TVEM	23
3.4	Data Analysis Example	23
3.4.1	Simulation study on estimating nonlinear coefficient functions	23
3.4.2	Simulation study mimicking real data	24
3.4.3	Practical concerns during data analysis	27
3.5	Extensions of TVEM	28
4	Discussions	29
A	Simulation codes for MELS	30
B	Simulation codes for mixed location-scale HMM	34

1 Introduction

Ecological momentary assessment (EMA) is a group of methods using real-time data capture techniques such as timed smartphone prompts to repeatedly sample subjects' behaviors, thoughts, and states in their natural environments. These methods have seen wide applications in the field of psychology, sociology, and public health sciences where there is an interest to study subjects in settings close to real world.¹

Traditional data collection methods such as retrospective self-reports suffer from recall bias, and are unsuited to capture momentary and contextual information needed to understand how subject behaviors and experiences unfold and vary over time.² EMA uses devices like smartphones to collect data at scheduled times (could be at fixed or random intervals) or in response to a clinically relevant prompt such as a smoking event. The repeated assessment of subjects reveals how a single subject's state varies across time and situations, allowing researchers to study within-person processes with higher degree of ecological validity.

Beyond academic research, EMA data are also integrated into mHealth (mobile health) platforms which collect data remotely and use them to provide well-timed, individually tailored interventions. Many mHealth platforms aim at treating conditions such as depression, dietary disorder, and substance abuse, all of which require day-to-day monitoring of and intervention on subjects' thoughts and behaviors. These data are hard to collect passively, and would usually require subjects' self-reports. Given this, EMA is essential as it is one of the few, if not the only way to collect high-resolution data on subjects' subjective experiences as they unfold in real life.

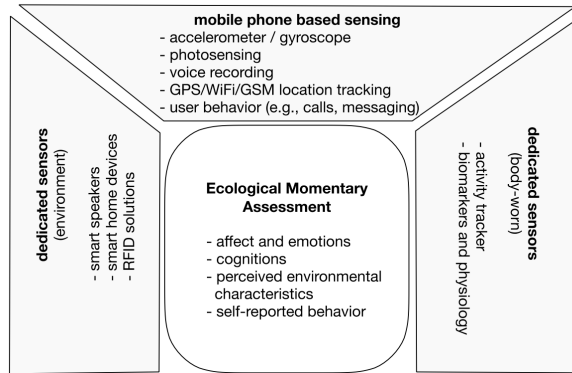


Figure 1: Ecological Momentary Assessment within a mobile sensing framework

As pictured above, when synchronized with continuous physiological data and contextual information collected from passive sensors such as GPS and wearable devices, EMA data provides substantial opportunities to model the dynamics of various clinical and psychological processes, as well as inferring ecological explanations for behaviors such as smoking and binge drinking.³

This paper examines two models that are often used in existing literature to take advantage of the rich temporal and contextual information embedded within EMA data. Mixed-effects location-scale model⁵ takes advantage of the intensive sampling of within-individual data over time and across contexts to jointly model the effects of covariates (usually some behavioral or contextual cues such as the act of smoking) on both the mean and variability of the response variable (usually mood, affect, or other measures of subject experience). Time-varying effects model¹⁶ takes advantage of the temporal resolution of EMA data to reveal changing nature of associations between key variables throughout some clinical process of interest, such as the changing relationship between smoking and mood throughout a smoking cessation attempt. We will motivate these models with studies, establish and verify their theoretical properties through simulations, and assess their strengths and weaknesses.

2 Mixed-effects location-scale model

2.1 A motivating study

The motivating study for MELS is the EMA portion of a longitudinal natural history study of adolescent smoking conducted by Mermelstein et al in 2002.⁴ For the study, 510 adolescents who were in 8th or 10th grade at the beginning of the study were recruited. The subjects were either non-smokers who indicated interest in smoking or light smokers who smoked at least once in the past 90 days before baseline but have not smoked more than 100 cigarettes in a lifetime. Subjects carried hand held computers with them over a seven-day data collection period, and were asked to respond to interview prompts appearing randomly on the device or to initiate a data collection interview themselves whenever they smoked. The interview asked range of subjective evaluation of mood items (I felt happy, I felt relaxed, I felt cheerful...etc) identified via factor analysis as well as questions about the place, activity, and presence of companionship at the time of response. The response variables of interest are positive affect level and negative affect level. Both measures were constructed by averaging over several mood items that were identified via factor analysis, and were rated from 1 to 10. The two measures are constructed separately: higher score for positive mood items indicates higher level of positive affect, and higher score for negative mood items indicates higher level of negative affect.

The interest is to see whether smoking increases or decreases the mean and variability of either positive or negative affects. In other words, in addition to examining whether smoking could elevate mood, we also want to examine whether it stabilizes or destabilizes mood. This is done by comparing data collected during smoking episodes against data collected from random prompts. Mixed-effects location-scale model⁵ is developed to jointly model the mean and variance of affect levels while carrying out correct inference for the effects of covariates on variance by accounting for individual heterogeneity in both mean (location) and variability (scale).

2.2 Formulation of MELS

The basis of MELS is linear mixed effect model,⁶ which is motivated by the nested structure of intensive longitudinal data whereby observations made on the same individual over time are correlated. Let y_{it} denote the measurement of response variable such as positive affect level for subject i ($i = 1, 2, \dots, N$ subjects) on occasion t ($t = 1, 2, \dots, T_i$ occasions), we can model y_{it} as:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \nu_i + \epsilon_{it} \quad (1)$$

\mathbf{x}_{it} is a p -dimensional vector of covariates including an intercept term. The covariates may be at the subject-level (level 2) such as gender, or at the occasion level (level 1) such as whether the subject was smoking or not during that particular occasion, or it could be an interaction of level 1 and level 2 covariates. The random location effect ν_i is assumed to be normally distributed with mean 0 with variance σ_ν^2 . The random effect ν_i represents individual i 's deviation from $\mathbf{x}_{it}'\boldsymbol{\beta}$, which is the expected value of response variable conditioning on covariates \mathbf{x}_{it} . The larger σ_ν is, the more heterogeneity there is in individuals' deviation from the expected value of response at any covariates value. We call this type of variation between-subject variance (BS-variance). The error term ϵ_{it} is assumed to be normally distributed with mean 0 and variance σ_ϵ^2 . They are assumed to be independent across subjects and observations. The error represents the deviation from expected value of response given covariates and individual random location effect, and the error variance σ_ϵ^2 represents the variability in this deviation. We call this variability the within-subject variance (WS-variance).

The main feature of MELS is that WS-variance is no longer considered homogeneous across subjects and covariate values. To model the WS-variance in terms of covariates, we can represent the variance parameter $\sigma_{\epsilon_{it}}^2$ using log-linear representation which guarantees a positive value no matter the value of regression coefficients:

$$\sigma_{\epsilon_{it}}^2 = \exp(\boldsymbol{\omega}_{it}'\boldsymbol{\tau}) \quad (2)$$

Similarly, we can also represent the BS-variance term using log-linear representation. For simplicity, we will first consider the BS-variance homogeneous across covariates values here, so we only have an intercept term:

$$\sigma_\nu^2 = \exp(\alpha_0) \quad (3)$$

ω_{it} is the vector of covariates thought to be associated to the WS-variance, with τ being the corresponding coefficients. The sets of covariates used in the WS-variance sub-model can be the same or different from the covariates used for the mean sub-model. The coefficients τ represent the effects of covariates on WS-variance in log-scale. If $\tau_p > 0$, then a one-unit increase in the p -th covariate increases the WS-variance by the multiplicative factor of $\exp(\omega'_{it}\tau + \tau_p)/\exp(\omega'_{it}\tau) = \exp(\tau_p)$. When reporting the results for MELS models, the coefficients for variance sub-models are usually reported after exponentiation. Therefore, a 95% confidence interval containing 1 indicate that the covariate has no statistically significant effect on variance at 5% significance level.

Just like how subjects can exert individual influences on mean affect level beyond the effects of covariates, they can also exert influence on the variance of affect level, and we model this by including a random scale effect ω_i in the model for WS-variance:

$$\sigma_{\epsilon_{it}}^2 = \exp(\omega'_{it}\tau + \omega_i) \quad (4)$$

Like the location effect ν_i , ω_i is assumed to follow a normal distribution with mean 0, and is allowed to co-vary with the location effect through the covariance parameter $\sigma_{\nu\omega}$:

$$\begin{bmatrix} \nu_i \\ \omega_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\nu^2 & \sigma_{\nu\omega} \\ \sigma_{\nu\omega} & \sigma_\omega^2 \end{bmatrix} \right) \quad (5)$$

The inclusion of ω_i gives the WS-variance a log-normal distribution:

$$\sigma_{\epsilon_{it}}^2 \sim \text{lognormal}(\omega'_{it}\tau, \sigma_\omega^2) \quad (6)$$

With the inclusion of both the location and scale effects, the conditional variance of measurement y_{it} given covariates ω_{it} is now:

$$\text{Var}(y_{it}|\omega_{it}) = \text{Var}(\nu_i) + \text{Var}(\epsilon_{it}) = \exp(\alpha_0) + \exp(\omega'_{it}\tau + \frac{1}{2}\sigma_\omega^2) \quad (7)$$

, where for the second term of the second step we took the expected value of the log-normal distribution followed by $\sigma_{\epsilon_{it}}^2$.

To better understand the different sources of variation in the model, we can look at the intraclass correlation r_{it} of observation y_{it} :

$$r_{it} = \frac{\exp(\alpha_0)}{\exp(\alpha_0) + \exp(\omega'_{it}\tau + \frac{1}{2}\sigma_\omega^2)} \quad (8)$$

The intraclass correlation (ICC) represents the proportion of unexplained variation by covariates that can be accounted for by subject-level variation. Because we allow WS-variance to differ across subjects-level and occasion-level covariates, this proportion is changing across i and t . For example, if smoking largely decreases the WS-variance, then the ICC for subjects when they are smoking will be much larger than the ICC when they are not smoking.

2.3 Estimation of MELS model

We can use either likelihood-based methods⁷ or Bayesian methods⁸ to estimate the MELS model. In either case, we need to deal with the distribution of the random effects as we are either integrating it out to get the marginal likelihood or using it as priors. Therefore, standardizing the random effects to have a independent standard bivariate normal distribution can make computations more efficient. We do this through the Cholesky decomposition, which finds a lower triangular matrix \mathbf{S} such that:

$$\begin{bmatrix} \nu_i \\ \omega_i \end{bmatrix} = \mathbf{S} \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} = \begin{bmatrix} s_1 & 0 \\ s_2 & s_3 \end{bmatrix} \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} = \begin{bmatrix} \sigma_\nu & 0 \\ \sigma_{\nu\omega}/\sigma_\nu & \sqrt{\sigma_\omega^2 - \frac{\sigma_{\nu\omega}^2}{\sigma_\nu^2}} \end{bmatrix} \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} \quad (9)$$

where θ_{1i} and θ_{2i} are independent standard normals. It is easily shown that the matrix \mathbf{S} is the Cholesky factor where:

$$\text{Cov} \left(\begin{bmatrix} \nu_i \\ \omega_i \end{bmatrix} \right) = \mathbf{S} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{S}^T = \mathbf{S} \mathbf{S}^T = \begin{bmatrix} \sigma_\nu^2 & \sigma_{\nu\omega} \\ \sigma_{\nu\omega} & \sigma_\omega^2 \end{bmatrix} = \Sigma_{\nu\omega} \quad (10)$$

Now, we can re-write the expressions for mean and WS-variance model using entries of the Cholesky factor:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + s_1\theta_{1i} + \epsilon_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \sigma_\nu\theta_{1i} + \epsilon_{it} \quad (11)$$

$$\sigma_{\epsilon_{it}}^2 = \exp(\boldsymbol{\omega}'_{it}\boldsymbol{\tau} + s_2\theta_{1i} + s_3\theta_{2i}) \quad (12)$$

With this, we estimate the parameters s_1, s_2 and s_3 instead, and the original variance-covariance parameters are recovered as $\sigma_{\nu\omega} = s_1 \times s_2$, $\sigma_\nu^2 = s_1^2$, and $\sigma_\omega^2 = s_2^2 + s_3^2$.

We can further simplify the model by forcing $\sigma_{\nu\omega} = 0$, so s_2 is also forced to be 0. Instead of modeling the effect of location effect on WS-variance through the covariance parameter $\sigma_{\nu\omega}$, we can include the standardized location effect θ_{1i} directly into the model for WS-variance as a covariate:

$$\sigma_{\epsilon_{it}}^2 = \exp(\boldsymbol{\omega}'_{it}\boldsymbol{\tau} + \tau_\ell\theta_{1i} + \tau_q\theta_{1i}^2 + \sigma_\omega\theta_{2i}) \quad (13)$$

Here, both the linear term θ_{1i} and the quadratic term θ_{1i}^2 are included to model a quadratic relationship between the random location effect and the log of WS-variance. Such quadratic relationship can be observed when there is ceiling or flooring effect in the data, where subjects with extreme baseline levels are much less variable than others. With this simpler formulation, we estimate the parameters $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, α_0 , and $\mathbf{s}' = [\tau_\ell, \tau_q, \sigma_\omega]$.

2.3.1 Maximum likelihood estimation

For maximum likelihood-based estimation, we first write out the distribution of the error:

$$\epsilon_{it} = y_{it} - (\mathbf{x}'_{it}\boldsymbol{\beta} + \sqrt{\exp(\alpha_0)}\theta_{1i}) \sim N(0, \sigma_{\epsilon_{it}}^2) \quad (14)$$

$$\sigma_{\epsilon_{it}}^2 = \exp(\boldsymbol{\omega}'_{it}\boldsymbol{\tau} + \mathbf{s}'\boldsymbol{\theta}_i), \text{ where } \boldsymbol{\theta}'_i = [\theta_{1i} \quad \theta_{2i}] \quad (15)$$

The conditional likelihood for subject i with T_i observations, conditioning on the values of standardized location and scale random effects $\boldsymbol{\theta}_i$ is:

$$\ell(\boldsymbol{\epsilon}_i|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\exp(\boldsymbol{\omega}'_{it}\boldsymbol{\tau} + \tau_\ell\theta_{1i} + \tau_q\theta_{1i}^2 + \sigma_\omega\theta_{2i})}} \exp \left(-\frac{(y_{it} - (\mathbf{x}'_{it}\boldsymbol{\beta} + \sqrt{\exp(\alpha_0)}\theta_{1i}))^2}{2\exp(\boldsymbol{\omega}'_{it}\boldsymbol{\tau} + \tau_\ell\theta_{1i} + \tau_q\theta_{1i}^2 + \sigma_\omega\theta_{2i})} \right) \quad (16)$$

and the marginal likelihood for subject i over the joint distribution of the two random effects is:

$$h(\boldsymbol{\epsilon}_i) = \int_{\boldsymbol{\theta}} \ell(\boldsymbol{\epsilon}_i|\boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta} \quad (17)$$

The log-likelihood for all N subjects is then:

$$\ln L = \sum_{i=1}^N \ln h(\boldsymbol{\epsilon}_i) \quad (18)$$

where $g(\boldsymbol{\theta})$ in equation (17) is the bivariate standard normal distribution. We maximize this log marginal likelihood to get the ML estimates for our parameters. Each subject's marginal likelihood $h(\boldsymbol{\epsilon}_i)$ is computed using numerical integration methods such as quadrature methods. The approximated integrals then feed into the computation of gradient and Hessian, which are used to update parameter estimates through Newton-Raphson iteration. The process is repeated until the algorithm converges, and the standard errors for ML estimates including z-statistics for Wald-test of significance are derived from the same large sample theory followed by linear mixed models.

Finally, we can estimate the random location and scale effects for each individual through Empirical Bayes methods. Denoting the conditional likelihood for subject i as ℓ_i , the joint posterior distribution for the random location and scale effects of subject i is:

$$p(\boldsymbol{\theta}_i|\boldsymbol{\epsilon}_i) = \frac{p(\boldsymbol{\epsilon}_i|\boldsymbol{\theta}_i) \cdot p(\boldsymbol{\theta}_i)}{p(\boldsymbol{\epsilon}_i)} = \frac{\ell_i \cdot g(\boldsymbol{\theta}_i)}{h(\boldsymbol{\epsilon}_i)} \quad (19)$$

We estimate the random effects by computing the expectation of this posterior distribution over the standard bivariate normal distribution. The conditional likelihood ℓ_i and marginal likelihood $h(\boldsymbol{\epsilon}_i)$ are computed using parameters estimated from maximum likelihood, hence making the method "Empirical" instead of fully Bayes:

$$\mathbb{E}(\boldsymbol{\theta}_i|\boldsymbol{\epsilon}_i) = \frac{1}{h(\hat{\boldsymbol{\epsilon}}_i)} \int_{\boldsymbol{\theta}} \boldsymbol{\theta} \hat{\ell}_i g(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (20)$$

There are several software for implementing the maximum likelihood estimations of MELS model. MIXREGLS⁷ is a stand-alone program that estimate the MELS model as formulated in this paper so far. SAS PROC NLXMIED can also be tweaked to accommodate the estimation of MELS models. MIXWILD¹⁰ is another stand-alone program, and it extends MIXREGLS's capacity by allowing for the inclusion of multiple location and scale random effects.

2.3.2 Bayesian approaches

Alternatively, we can use Bayesian approaches to estimate the model parameters which has been shown to produce estimates very similar to maximum likelihood estimates with uninformative priors.⁸ In Bayesian analysis, all the model parameters estimated in ML method ($\boldsymbol{\beta}, \boldsymbol{\tau}, \alpha_0, \mathbf{s}$) as well as the random effects θ_{1i}, θ_{2i} are treated as random with some prior distributions. To minimize the influence of priors on estimates, we usually choose very diffuse priors such as the improper uniform distribution (uniform on $(-\infty, +\infty)$) over regression coefficients ($\boldsymbol{\beta}, \boldsymbol{\tau}, \alpha_0, \tau_\ell, \tau_q$), and improper bounded uniform distribution (uniform on $(0, +\infty)$) for variance parameter (σ_ω). The priors for the random effects are standard bivariate normals by construction. Let \mathbf{Y}_{it} be the matrix of response variable for all subjects at all occasions. $P(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}, \alpha_0, \mathbf{s}|\mathbf{Y}_{it})$ is the joint posterior given the data, and we will approximate this joint posterior distribution by drawing from the conditional posteriors for each group of parameters. For instance, the conditional posterior for fixed effects $\boldsymbol{\beta}$ in the mean model would be:

$$P(\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\tau}, \alpha_0, \mathbf{s}, \mathbf{Y}_{it}) \propto \mathbf{P}(\mathbf{Y}_{it}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}, \alpha_0, \mathbf{s})\pi(\boldsymbol{\beta}) \quad (21)$$

with $\pi(\cdot)$ denoting the prior distribution for each parameter vectors. Other conditional posteriors have similar forms. The conditional likelihood $P(\mathbf{Y}_{it}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}, \alpha_0, \mathbf{s})$ is the likelihood used in ML estimate in equation (16) multiplied together over all N subjects. Iterative sampling from all the conditional posteriors are done through common Bayesian algorithms such as Metropolis-Hastings-Gibbs sampling algorithm or Hamiltonian Monte-Carlo methods.

While the choice of estimation methods is usually a matter of convenience and software availability, when we want to include many more random effects such as random slopes or another level of random effects, maximum likelihood methods could face problems in integrating over the joint-distribution of numerous random effects. Bayesian methods on the other hand are less affected by the number of random effects in the model, making them the more appropriate method.

2.4 Simulation study of MELS

2.4.1 Impact of scale effect on inference

The main feature of MELS is the inclusion of random scale effect in the joint-modeling of mean and variance. To illustrate the importance of the random scale effect, we compare the performance of model with and without it in a simulation study.

Due to limitations of computation resources, we will limit the number of our data sets to 50 for all subsequent simulations. For each data set, we simulated 150 subjects measured at 25 occasions. We generate one binary between-subject covariate $X_i^{BS} \sim \text{Bernouli}(0.5)$ and one continuous within-subject covariate $X_{it}^{WS} \sim N(0,1)$. These two covariates were used in both the mean and WS-variance sub-models. The covariance $\sigma_{\nu\omega}$ between random location and scale effects is set to 0, and we instead allow the location effect to be log-linearly related to the WS-variance through parameter τ_ℓ . The true parameter values are $\beta' = [1, 1, 1]$, $\tau' = [0.5, 0.6, -0.4]$, $\sigma_\nu = 1$, $\sigma_\omega = 0.8$, and $\tau_\ell = 0.3$.

For each data set, we fit models with and without the random scale effect. Because the linear effect of location effect on WS-variance is introduced as an alternative way to model the covariance between location and scale effects, we also exclude location effect term from the WS-variance sub-model for the model without scale effect.

Model Without Scale Effect:

$$y_{it} = \beta_{intercept} + \beta_{BS}x_i^{BS} + \beta_{WS}x_{it}^{WS} + \nu_i + \epsilon_{it} \quad (22)$$

$$\sigma_{\epsilon, it}^2 = \exp(\tau_{intercept} + \tau_{BS}x_i^{BS} + \tau_{WS}x_{it}^{WS}) \quad (23)$$

Model With Scale Effect:

$$y_{it} = \beta_{intercept} + \beta_{BS}x_i^{BS} + \beta_{WS}x_{it}^{WS} + \nu_i + \epsilon_{it} \quad (24)$$

$$\sigma_{\epsilon, it}^2 = \exp(\tau_{intercept} + \tau_{BS}x_i^{BS} + \tau_{WS}x_{it}^{WS} + \tau_\ell\theta_{1i} + \sigma_\omega\theta_{2i}) \quad (25)$$

The models are estimated using Bayesian methods, implemented through Stan which uses Hamiltonian Monte-Carlo algorithm. Diffuse priors are selected for all parameters, as specified in the previous section on Bayesian estimation. The results are presented in Table 1 next page. We report the average interval widths of 95% credible interval constructed from taking quantile values of the posterior. We use the term "coverage" loosely here as the interpretation of Bayesian credible interval is different from frequentist confidence interval. However, with sufficient sample and appropriate prior, they usually exhibit similar behaviors.⁹

Table 1: Model Estimates for 50 Simulated Data Sets

Parameter	Without Scale Effect			With Scale Effect		
	Bias	AIW	Coverage	Bias	AIW	Coverage
$\beta_{intercept} = 1$	0.018	0.482	98%	0.016	0.482	98%
$\beta_{BS} = 1$	-0.054	0.686	92%	-0.055	0.689	92%
$\beta_{WS} = 1$	-0.003	0.111	94%	0.001	0.081	96%
$\tau_{intercept} = 0.5$	0.376	0.131	0%	0.014	0.414	96%
$\tau_{BS} = 0.6$	-0.029	0.185	34%	-0.003	0.589	100%
$\tau_{WS} = -0.4$	0.008	0.094	84%	0.005	0.099	94%
$\sigma_\nu = 1$	0.012	0.260	96%	0.016	0.260	98%
$\sigma_\omega = 0.8$	-	-	-	0.002	0.215	96%
$\tau_\ell = 0.3$	-	-	-	0.000	0.308	98%

From the results above, we see that both models give good estimates for β parameters in the mean sub-model. However, for the model without scale effect, the estimates of WS-variance coefficients are biased

and give insufficient coverage. The intercept for WS-variance sub-model is biased upward and provides an abysmal 0% coverage. By not including the random scale effect when there actually is individual differences in baseline variability of the response variable, we are ignoring a major source of variation. Consequently, the intercept $\tau_{intercept}$ is biased upward to account for the ignored source of uncertainty. The coverage for $\tau_{intercept}$ and τ_{BS} are both insufficient, because we underestimate the uncertainty involved in the estimation of these parameters by ignoring the subject-level heterogeneity in variability of response variable. Therefore, not including the random scale effect can lead to biased and over-confident estimation of parameters in the WS-variance sub-model, resulting in false positive findings when we want to understand the effects of covariates on the consistency of response variable.

2.4.2 Robustness to misspecification of covariance structure

So far in the model, we have assumed that the error terms ϵ_{it} are independent across subjects and across all observations for each subject. However, it is sensible to suspect some sort of correlation between temporally contiguous observations. We may be tempted to simply include lagged dependent variables $y_{i,t-1}, y_{i,t-2} \dots$'s as covariates for the mean and variance sub-models of $y_{i,t}$, but this would violate our assumption that random effects ν_i 's and ω_i 's are independent from covariates in the mean and WS-variance sub-models, and would result in invalid inference. Instead, we can employ AR(1) correlation structure proposed by Chi and Reinsel.¹¹ In this model, we specify a $T_i \times T_i$ correlation matrix P_i for each subject i , where ρ is the correlation coefficient between 0 and 1:

$$P_i = \begin{bmatrix} 1 & \rho & \dots & \rho^{T_i-1} \\ \rho & 1 & \dots & \rho^{T_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T_i-1} & \rho^{T_i-2} & \dots & 1 \end{bmatrix} \quad (26)$$

Here, we assume that the indices of observations within each subject correspond to the temporal order at which they are collected. When detailed time-stamp information is available, we can also modify the correlation matrix so that correlation between two assessments decays exponentially over some distance measure based on time stamp. The within-subject variance-covariance matrix Σ_{ϵ_i} for each subject i is then:

$$\Sigma_{\epsilon_i} = \begin{bmatrix} \sigma_{\epsilon_{i,1}}^2 & \sigma_{\epsilon_{i,1}}\sigma_{\epsilon_{i,2}} & \dots & \sigma_{\epsilon_{i,1}}\sigma_{\epsilon_{i,T_i}} \\ \sigma_{\epsilon_{i,2}}\sigma_{\epsilon_{i,1}} & \sigma_{\epsilon_{i,2}}^2 & \dots & \sigma_{\epsilon_{i,2}}\sigma_{\epsilon_{i,T_i}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\epsilon_{i,T_i}}\sigma_{\epsilon_{i,1}} & \sigma_{\epsilon_{i,T_i}}\sigma_{\epsilon_{i,2}} & \dots & \sigma_{\epsilon_{i,T_i}}^2 \end{bmatrix} \odot \begin{bmatrix} 1 & \rho & \dots & \rho^{T_i-1} \\ \rho & 1 & \dots & \rho^{T_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T_i-1} & \rho^{T_i-2} & \dots & 1 \end{bmatrix} \quad (27)$$

Standard software for estimating MELS such as MIXREGLS and MIXWILD mentioned previously currently do not have the option to specify AR(1) structure of correlation. While searching for an efficient way to implement the computation of this model is a topic of future research, we conduct a simulation study to examine the consequences of estimating MELS model when AR(1) correlation of error terms exists in data but is ignored. For this simulation study, we use the same data generating mechanism as our previous simulation, with the addition of AR(1) correlation for observations made on the same individual. We simulate on setting of weak and strong autocorrelation, setting $\rho = 0.3$ and $\rho = 0.7$ respectively. We simulated 50 data sets for each setting and estimated model parameters using Bayesian method implemented in Stan as our previous simulation. The result is produced in Table 2 on next page.

We can see that the estimation of fixed effects for the mean-submodel is robust, with low bias and coverage rate near 95% for both low and high autocorrelation cases. However, when auto-correlation is high, estimates and inference for the WS-variance sub-model are no longer valid as we have completely ignored the AR(1) correlation structure of errors on observations nested in a subject. The strong auto-correlation would make level-1 errors look more homogeneous across different covariate values than they actually are, which results in the estimates of covariates τ_{BS} and τ_{WS} to be biased towards 0 and $\tau_{intercept}$ heavily biased upwards.

Table 2: Model estimates for 50 simulated data sets

Parameter	$\rho = 0.3$			$\rho = 0.7$		
	Bias	AIW	Coverage	Bias	AIW	Coverage
$\beta_{\text{Intercept}} t = 1$	0.014	0.498	94%	-0.002	0.613	96%
$\beta_{\text{BS}} = 1$	-0.039	0.711	94%	-0.042	0.864	98%
$\beta_{\text{WS}} = 1$	-0.003	0.079	92%	-0.004	0.125	94%
$\tau_{\text{intercept}} = 0.5$	-0.029	0.419	96%	1.142	0.456	0%
$\tau_{\text{BS}} = 0.6$	-0.001	0.59	98%	-0.568	0.645	0%
$\tau_{\text{WS}} = -0.4$	-0.009	0.099	92%	0.407	0.103	0%
$\sigma_{\nu} = 1$	0.056	0.269	90%	0.235	0.348	16%
$\sigma_{\omega} = 0.8$	0.0102	0.216	92%	0.147	0.241	32%
$\ell = 0.3$	-0.001	0.309	98%	-0.287	0.37	18%

The estimates of BS-variances ($\sigma_{\nu}, \sigma_{\omega}$) are also biased upwards, potentially to account for the auto-correlation that exist for observations within the same subject. This results in a misleading intraclass-correlation-coefficient (ICC), whereby more variations are attributed to between-subject heterogeneity when it's in fact coming from auto-correlation within-subject. While the impact of high auto-correlation is pronounced, when auto-correlation is low, estimates and inferences for the WS-variance sub-model parameters aren't affected, as both bias and coverage are good, showing that MELS model is robust to the presence of low auto-correlation.

In practice, the serial dependence of observations could be even more complicated, as EMA data are collected over several days, so observations could be nested within days which are in turn nested within subjects. It might not make sense to include the last observation in a day in the model for the first observation of the next day, so we would have separate covariance matrix for observations made within each day. Furthermore, there could be inter-day correlations whereby the mean-level of response between days could be correlated due to subjects' lifestyle and personality traits. While there are various models that only model the mean of response variable while accounting for complicated error correlation structure,¹² currently there are no widely used models that can jointly model mean and variance of response variable while taking into account complex correlation structure.

2.5 Mixed location-scale hidden Markov model

2.5.1 Motivation and model formulation

As demonstrated in the simulation study above, MELS is not robust to misspecification of WS-covariance structure, and current software for running MELS models cannot easily estimate models with correlated error structure. One way to deal with this is to include covariates that could account for serial correlation in observations. For instance, level-1 covariates such as "being with others" may have residual effects that last across observations, so including lagged terms of these covariates could partially account for the auto-correlation in errors and bring it down to a level that can be ignored.

An alternative way to deal with auto-correlation is to employ mixed location-scale hidden Markov model (HMM), where serial-dependence of observations is instead modelled by subject's tendencies to stay in some discrete latent states.¹³ HMM is not exactly a model that deals with data generated under AR(1) correlation structure as formulated in equation (26). Instead, it makes the assumption that subjects go through a sequence of latent, discrete states (emotional, cognitive, or behavioral states) over the observation period, and this sequence is modelled by a first-order Markov chain. Furthermore, observed response variable (which we also call observed process) at each observation point is modelled conditioning on the latent state of subject at that time point. While not all psychological and clinical processes studied using EMA lend itself to such assumption, when they do, mixed location-scale HMM is a powerful tool to model the different levels of variability in EMA data while accounting for serial dependence of observations. In particular, between-subject variability is accounted for through the inclusion of location and scale random effects, while serial

dependence of within-subject observations are modelled by transition probabilities across latent states. In our case with the motivating study on adolescent smoking, we may posit that there are different affective states adolescents can fall into, and across different states, adolescents could respond differently to smoking and being alone with regard to their mood regulation.

The model formulations is as follows. Let there be $i = 1, 2, \dots, N$ subjects, each with T_i observations. We further assume that each observations are taken over approximately equally spaced time intervals, which is essential for discrete HMM. This assumption could be relaxed if we employ continuous-time domain Markov model, but for the purpose of this paper we will limit our model to discrete Markov models.[?]

For simplicity, we assume that there are only two discrete latent states, state 1 and state 2. Then for individual i with T_i observations, we define $Z_i = (z_{i,1}, z_{i,2}, \dots, z_{i,T_i})$ to be an unobserved binary random vector whose elements follow a first order Markov chain with unknown transition probability matrix A :

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad a_{11} + a_{12} = 1, \quad a_{21} + a_{22} = 1 \quad (28)$$

and initial probability distribution $(\pi_1, \pi_2 = 1 - \pi_1)$. The first-order Markov assumption then implies that:

$$P(z_{i,1} = k) = \pi_k, \quad k = 1 \text{ or } 2 \quad (29)$$

$$P(z_{i,t} | z_{i,1}, \dots, z_{i,t-1}) = P(z_{i,t} | z_{i,t-1}) = a_{z_{i,t-1} z_{i,t}} \quad (30)$$

While it's possible to include random effects into the transition probability to reflect individual heterogeneity in the hidden process, this would make the estimation of HMM prohibitively complex and intractable. For simplicity, we will only model individual heterogeneity in the observed process as below:

$$y_{it} | z_{it} = k, \mathbf{x}'_{it} \boldsymbol{\beta}_k, \mathbf{w}'_{it} \boldsymbol{\alpha}_k, \nu_i, \omega_i \sim \mathcal{N}(\mathbf{x}'_{it} \boldsymbol{\beta}_k + \nu_i, \exp(\mathbf{w}'_{it} \boldsymbol{\alpha}_k + \omega_i)) \quad (31)$$

Here, z_{it} is the unknown latent state subject i is in at time t , which is assumed to be some state k here. \mathbf{x}_{it} and \mathbf{w}_{it} are the covariates for mean and WS-variance sub-models respectively, and their corresponding coefficient parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are different across latent states, so they are sub-scripted by the latent state k . We include both location and scale random effects ν_i and ω_i , which follows a bivariate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma_{\nu, \omega})$ as specified in equation (5) from the MELS section. The parameters to be estimated are: $\boldsymbol{\theta} = (A, \pi_1, \boldsymbol{\beta}_k, \boldsymbol{\alpha}_k \text{ for } k = 1, 2)$, as well as the random effects (ν_i, ω_i) for each subject.

2.5.2 Estimation algorithm

There are many algorithms for estimating parameters in HMM, as well as for recovering the most probable sequence of hidden states. Here, we will only attempt to estimate the parameters in HMM and conduct inference on our estimates. To deal with complexities introduced by random effects, we employ the forward-algorithm using Bayesian approach. In particular, we specify priors on all the parameters to be estimated: $(\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\omega})$, and run MCMC to sample from the posterior. The form of data likelihood used in Bayesian estimation is derived from the forwards algorithm, which consists of series of predict-update cycle computing filtered belief state $P(z_{i,t} = k | y_{i,1:t})$ for each subject i over each latent state k at each time point t . The computation is done online, meaning at each time point t , only information up to time point t is used. In the end, we will use the filtered belief states to compute the log likelihood of the data for each subject i : $P(y_{i,1:T_i} | \boldsymbol{\theta}, \omega_i, \nu_i)$

In the following, we will compute the data likelihood for subject i . In the prediction step, we compute the one-step-ahead predictive density:

$$P(z_{it} = k | y_{i,1:t-1}) = \begin{cases} \pi_k, & \text{if } t = 1 \\ \sum_j P(z_{i,t} = k | z_{i,t-1} = j) P(z_{i,t-1} = j | y_{i,1:t-1}), & \text{otherwise} \end{cases} \quad (32)$$

We then use the one-step-ahead predictive density as the prior to compute the filtered belief state at time t in the update step using Bayes rule:

$$P(z_{i,t} = k | y_{i,1:t}) = \frac{1}{F_{i,t}} P(y_{i,t} | z_{i,t} = k, \boldsymbol{\theta}, \nu_i, \omega_i) P(z_{i,t} = k | y_{i,1:t-1}) \quad (33)$$

where the normalization constant $F_{i,t}$ is:

$$F_{i,t} = P(y_{i,t}|y_{i,1:t-1}) = \sum_j P(z_{i,t} = j|y_{i,1:t-1})P(y_{i,t}|z_{i,t} = j, \boldsymbol{\theta}, \nu_i, \omega_i) \quad (34)$$

We can then write the log likelihood of the data and use it for Bayesian estimation of model parameters:

$$\log P(y_{i,1:T_i}|\boldsymbol{\theta}, \nu_i, \omega_i) = \sum_{t=1}^{T_i} \log P(y_{i,t}|y_{i,1:t-1}) = \sum_{t=1}^{T_i} \log F_{i,t} \quad (35)$$

2.5.3 Simulation study

To verify the validity of our estimation procedure as well as examine the properties of mixed location-scale HMM, we conduct a simulation study. Data were generated according to the assumption of location-scale HMM, where individual heterogeneity across both mean and WS-variance exist. In particular, we generated our data as follows:

1. Assume there are only two states: state 1 or state 2. For $t = 1$, set probability of being in state 1 to $\pi_1 = 0.5$. For $t \geq 2$, generate the transition matrix where the probability of staying in state 1 is $a_{11} = 0.9$, and the probability of staying in state 2 is $a_{22} = 0.7$. With these probabilities, generate sequence of hidden states \mathbf{Z}_i for each subject i
2. Generate random subject location and scale effects according to bivariate normal distribution, where the means are 0, and $\sigma_\nu = 1$, $\sigma_\omega = \sqrt{0.3}$, and $\sigma_{\omega\nu} = -0.3\sqrt{3}$
3. For conditional distribution of response variable y , assume $y_{i,t}|z_{i,t} \sim \mathcal{N}(\mu_k + \nu_i, \sigma_{\epsilon,k} \sqrt{\exp(\omega_i)})$
4. Set $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_{\epsilon,1} = 2$, and $\sigma_{\epsilon,2} = 1$

We fit a regular HMM with no random effects, a mixed-effect HMM with only random location effect, and our proposed mixed location-scale HMM, and observe the following result:

Table 3: Model estimated for 50 data sets

Parameter	HMM			Mixed HMM			Mixed location-scale HMM		
	Bias	AIW	Coverage	Bias	AIW	Coverage	Bias	AIW	Coverage
$\pi_0 = 0.5$	-0.051	0.353	82%	-0.172	0.369	54%	-0.106	0.656	95%
$a_{11} = 0.9$	0.066	0.081	34%	-0.012	0.216	96%	-0.146	0.578	94%
$a_{22} = 0.7$	0.245	0.093	0%	0.181	0.174	16%	-0.037	0.589	97%
$\mu_1 = 1$	-0.662	0.765	14%	-0.323	0.954	72%	-0.253	1.479	94%
$\mu_2 = 2$	0.287	0.581	46%	-0.238	0.622	62%	-0.099	0.934	92%
$\sigma_{\epsilon,1} = 2$	0.314	0.421	24%	0.573	0.643	0%	0.069	0.854	96%
$\sigma_{\epsilon,2} = 1$	0.582	0.37	2%	0.219	0.399	40%	0.161	0.925	93%
$\sigma_\nu = 1$	-	-	-	-0.023	0.389	94%	0.007	0.429	93%
$\sigma_\omega = 0.548$	-	-	-	-	-	-	0.012	0.473	92%

When there are both location and scale random effects in the data generating process, mixed location-scale HMM is the only model that provides sufficient coverage and valid estimates. Mixed location-scale HMM successfully identifies the low-mean, high-variance state and the high-mean, low-variance state. For mixed HMM which ignores scale random effect, we see that the estimates for $\sigma_{\epsilon,1}$ and $\sigma_{\epsilon,2}$ are more separated than they actually are. By ignoring subject heterogeneity in WS-variance, mixed HMM attribute more of the total variation to differences in variance of observed process across states, when it should be explained by subject heterogeneity in WS-variance. Likewise, comparing HMM to the other two models that include a location random effect, we see that the estimates for μ_1 and μ_2 are more separated in HMM, as subject specific differences are being incorrectly attributed to differences in means across states.

While not implemented here, another analysis we can do is to use Viterbi algorithm to recover the most probable sequence of latent states for each subject. With this, we can further classify individuals based on the dominant state in which they fall into to gain additional insights on within-subject processes for each individual.

2.6 Data analysis example

To illustrate how MELS work in practice, we will analyze the data from the adolescent EMA study which we used as the motivating example for MELS. As only the random prompts portion of the data is publicly available, we will limit our analysis to that portion of the data. There were 17,402 random prompts obtained from 510 students, averaging 34 prompts per student (range = 3 to 58). Our interest is in understanding how covariates influence the mean and variance of positive and negative affect (PA and NA). We model positive and negative affects separately as studies have shown that they are distinct constructs instead of perfectly correlated opposite ends of a single spectrum.¹⁵ As negative affect and positive affect are modelled similarly, we will only model negative affect here. Although we have information on the day of time in which observations are made, timestamps of the temporal order of observations are not available so we cannot fit a AR(1) correlation structure model or hidden Markov model. We will therefore limit our analysis to implementing the simple MELS model, and examine the range of questions and analysis that can be done.

To justify the inclusion of random location and scale effects for both the mean and WS-variance sub-models, we make histograms of subject-level means and variances of PA and NA. Variances are represented in log-scale to correspond to the scale at which we are modeling them:

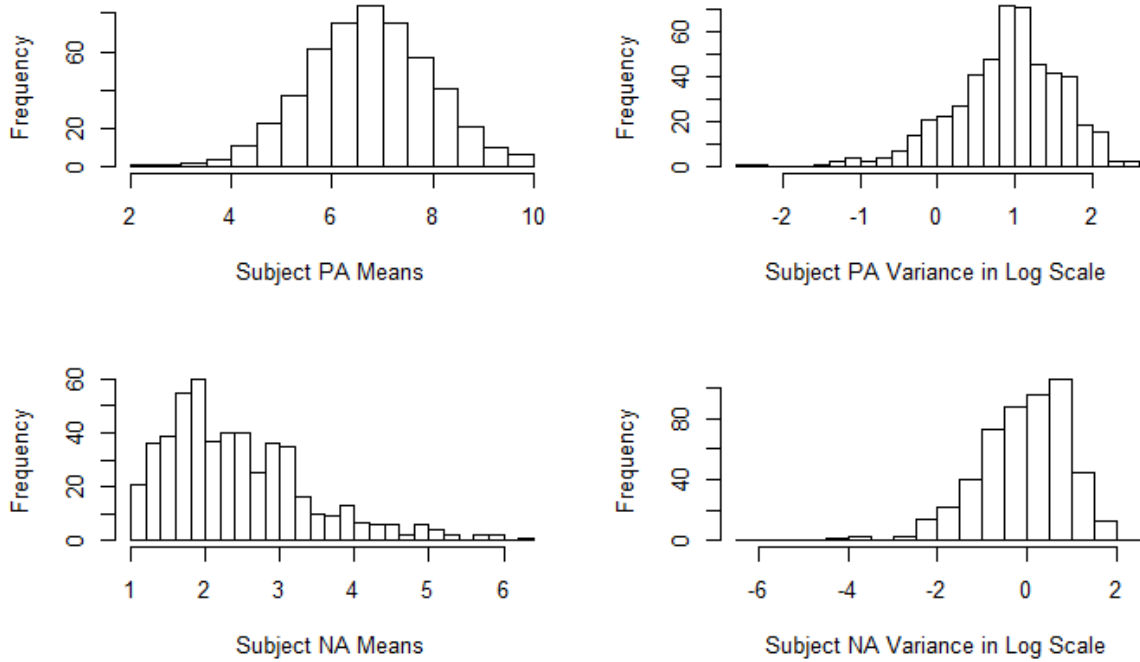


Figure 2: Histogram of Subject Means and Standard Deviations for PA and NA

From the histogram, we see that there is considerable heterogeneity in mean and variance of both positive affect and negative affect, which we account for by the inclusion of random location and scale effects. Modeling BS-variance corresponds to modeling the variance of the distribution of subject means represented by the histograms on left column. Modeling WS-variance corresponds to modeling the individual data points that make up the histograms on right column. Finally, estimating the variance of random scale effect corresponds to estimating the variance of the distributions represented by the histograms on right.

2.6.1 Null model for negative affect

We start our analysis by fitting an intercept-only model for NegAff, assuming quadratic relationship between the location and scale effect:

$$\text{NegAff}_{it} = \beta_0 + \sigma_\nu \theta_{1i} + \epsilon_{it} \text{ where } \theta_{1i} \sim N(0, 1) \quad (36)$$

$$\sigma_{\nu_{it}}^2 = \exp(\alpha_0) \quad (37)$$

$$\sigma_{\epsilon_{it}}^2 = \exp(\tau_0 + \tau_\ell \theta_{1i} + \tau_q \theta_{1i}^2 + \sigma_\omega \theta_{2i}) \text{ where } \theta_{2i} \sim N(0, 1) \quad (38)$$

The estimates are shown in the table below:

Table 4: *Intercept Only Mixed-Effects Location Scale Model of Negative Affect, N=510, $\sum T_i = 17,402$, maximum likelihood estimates*

Variable	Estimate	AsymStdErr	z.value	p.value
β_0	2.431	0.041	59.871	0.000
α_0	-0.213	0.066	-3.210	0.002
τ_0	0.243	0.058	4.161	0.000
τ_ℓ	0.979	0.056	17.512	0.000
τ_q	-0.311	0.031	-9.944	0.000
σ_ω	0.650	0.026	24.937	0.000

The intraclass correlation estimated by this model is:

$$r_{it} = \frac{\exp(-0.213)}{\exp(-0.213) + \exp(0.243 + \frac{1}{2}(0.979^2 + (-0.311)^2 + 0.650^2))} = 0.232 \quad (39)$$

indicating that approximately 23% of variation in negative affect is explained by subject heterogeneity. Strictly speaking, the ICC here should not be sub-scripted as it is constant across covariates at both subject and occasion level. Another analysis we can do at this point is to examine the shape of the plot of random location and scale effects, which are predicted using Empirical Bayes methods. The graph produced on next page shows a quadratic pattern which is verified by the significant linear and quadratic location effect terms on WS-variance in the model. Subjects on the bottom left have both low location and scale effects, indicating that they have low baseline negative affect and are also more emotionally stable. Subjects on top right have high baseline negative affect, and are also more emotionally erratic.

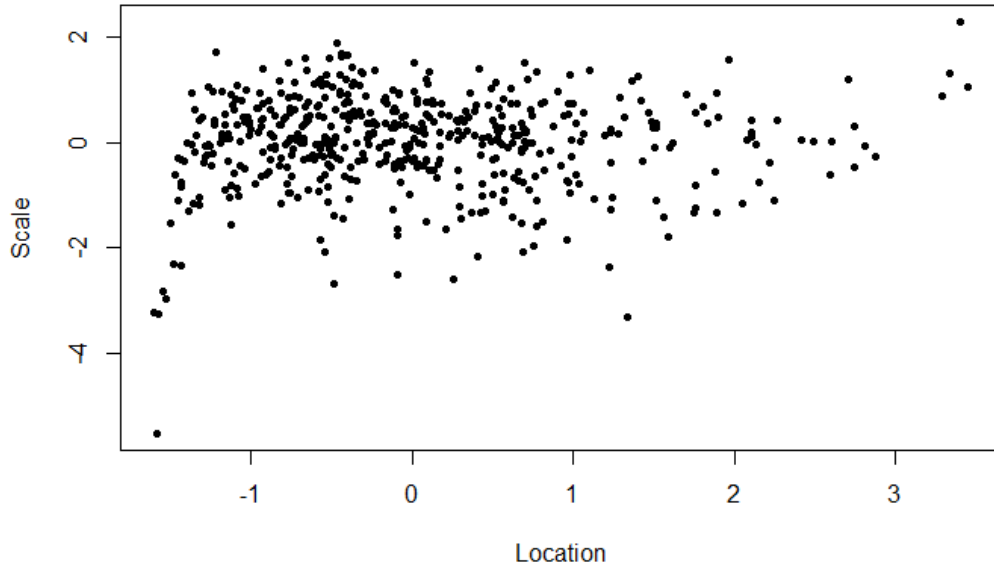


Figure 3: Plot of Location and Scale Effects for Intercept Only Model of Negative Affect

2.6.2 Modeling of BS-variance

Before moving on to fit the full-model, we can consider adding an extension to MELS model for this data set by modelling BS-variance σ_{ν}^2 in terms of both level-1 and level-2 covariates in manner similar to our model for WS-variance. That is to say, subjects from certain sub-populations with certain covariate values may be more heterogeneous compared to others. The addition is easy to implement in our current model with only a random location intercept. We represent the BS-variance using log-linear representation with covariates \mathbf{u}_{it} and corresponding coefficients $\boldsymbol{\alpha}$:

$$\sigma_{\nu_{it}}^2 = \exp(\mathbf{u}_{it}'\boldsymbol{\alpha}) \quad (40)$$

Model estimation does not change except that we replace all instances where $\sigma_{\nu_{it}}^2$ appears with the above log-linear representation, and there are additional entries for the $\boldsymbol{\alpha}$ terms in gradient and Hessian computation in Maximum Likelihood estimation.

The covariates \mathbf{u}_{it} can be either at subject level or occasion level. Effects of subject-level covariates such as gender are easy to understand; for instance, we might observe that boys are more similar to each other than girls are for the particular response variable, which implies a negative effect for a binary variable coded 1 for boys and 0 for girls. For occasion level covariates, consider a binary occasion level covariate such as the presence of companionship at time of response. At any occasion, this may be true (=1) or false (=0). This covariate can have effect on both the Within-Subject and Between-Subject variance as shown in the figure below:

There are two subjects in the figure above, each represented by red and blue. The WS-variance is represented by the dispersion of dotted observations around each subject's mean level, and the BS-variance is represented by the width of the gap between the mean levels of the two subjects. Here, being with others has a positive effect on the mean, but a negative effect on both the WS-variance and BS-variance.

2.6.3 Full model for negative affect

Moving on from the null model, we add several covariates to the mean, BS-variance, and WS-variance sub-model: *gender_f* is a subject level binary variable coded 0 for male and 1 for female; *Age15* is the subject's

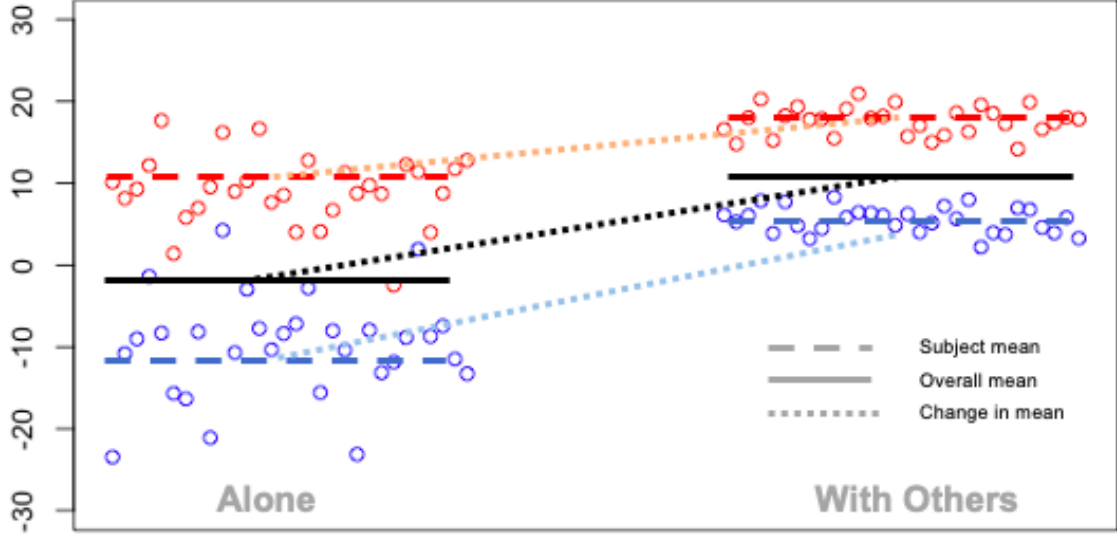


Figure 4: Effect of being with others on the mean, BS and WS variance

age minus 15, which is the average age of all subjects; *COPEc* is the grand-mean centered version of a coping scale measurement at baseline; $t1, t2, t3, t4$ are indicator variables coded to represent the time block (9am-2pm, 2pm-6pm, 6pm-10pm, and 10pm-3am) in which measurements are made, with the early morning period of 3am-9am used as reference level; finally, *Others* is a occasion level indicator variable coded 0 if subject was alone and 1 if subject was with others at the time of measurement. *Others* is decomposed to a within-subject and between-subject version (*Others_WS*, *Others_BS*). In particular, let \bar{Others}_i denote the proportion of time subject i is with others, then $Others_BS_i = \bar{Others}_i$, $Others_WS_{it} = Others_{it} - \bar{Others}_i$, and $Others_{it} = Others_WS + Others_BS$. The coefficient for *Others_BS* reveals how subject tendency to be with others influence negative affect, while the coefficient for *Others_WS* reveals how a momentary state of being with others influence negative affect.

The model estimates are reported in Table 3. Starting with the coefficients for the mean sub-model, we see that all time indicators have positive effects, with the first three (9am-2pm, 2pm-6pm, 6pm-10pm) being statistically significant at 5% significance level. This indicates that negative affect is elevated after the reference early morning period (3am-9am). Being female or being older than 15 years old do not have significant effect on negative affect, but having a higher coping score leads to lower negative affect. For *Others*, both the WS and BS versions have statistically significant negative effects. This means that a higher overall tendency to be with others as well as the momentary effect of being with others can both dampen negative affect. Moving on to the BS-variance sub-model, only *Others_WS* and *COPEc* are statistically significant, and both have negative effects. Therefore, when subjects are with others, there are less heterogeneity between subjects. There are also less heterogeneity for subjects with higher coping score. For the WS-variance sub-model, all coefficients are significant except for age. Negative affect level is more erratic for all time periods relative to the 3am-9am reference period, and females have more variable negative affect level. On the other hand, both WS and BS versions of *Others* and coping score makes negative affect level more consistent. The random scale standard deviation σ_w is significant, indicating there is indeed subject heterogeneity in the degree of baseline consistency of negative affect beyond what's explained by covariates alone. Finally, the random location effect has significant linear and quadratic effects on the WS-variance. The linear effect is positive, indicating that subjects with higher average negative affect is more erratic as well. The quadratic effect indicates that the rate at which WS-variance increase with location effect decelerates with higher values of location effect. The relationship also suggests the presence of floor effect, whereby subjects with low mean negative affect level are also consistently low in negative affect.

Table 4 below shows the exponentiated values of the coefficients in BS and WS-variance sub-models. The exponentiated coefficients represent the ratio of variance per unit increase in the corresponding regressor, so a 95% confidence level that does not include 1 indicates statistical significance. What's notable in Table 4 is that the 95% confidence for *Others_BS* in both BS and WS-variance sub-model is unusually wide (0.961 for BS sub-model and 0.65 for WS sub-model). This is because the coefficient represents the effect of a unit change in the covariate, but *Others_BS*, which represents a proportion, ranges only from 0 to 1. Therefore, the effect essentially represents the jump from one extreme (always alone) to another (never alone), leading to higher uncertainty in estimation and therefore wider CI.

Table 5: *Mixed-effects location scale model of negative affect estimated with maximum likelihood*

Variable	Estimate	Std Error	z-value	p-value
<i>β :Regression Coefficients</i>				
Intercept	2.721	0.143	19.037	0.000
9am-2pm	0.062	0.024	2.526	0.012
2pm-6pm	0.070	0.025	2.766	0.006
6pm-10pm	0.051	0.025	2.022	0.043
10pm-3am	0.029	0.041	0.709	0.478
Others_WS	-0.232	0.021	-11.005	0.000
Others_BS	-0.513	0.186	-2.766	0.006
genderf	0.008	0.057	0.134	0.894
age15	0.021	0.026	0.802	0.422
COPEc	-0.418	0.053	-7.869	0.000
<i>α :BS Variance parameters: log-linear model</i>				
Intercept	-0.164	0.279	-0.589	0.556
9am-2pm	0.077	0.052	1.463	0.143
2pm-6pm	0.102	0.054	1.899	0.058
6pm-10pm	0.015	0.055	0.277	0.781
10pm-3am	0.057	0.087	0.655	0.512
Others_WS	-0.303	0.042	-7.205	0.000
Others_BS	-0.531	0.380	-1.395	0.163
genderf	0.156	0.122	1.278	0.201
age15	0.078	0.053	1.470	0.141
COPEc	-0.833	0.119	-6.974	0.000
<i>τ :WS variance parameters: log-linear model</i>				
Intercept	0.300	0.222	1.350	0.177
9am-2pm	0.175	0.042	4.130	0.000
2pm-6pm	0.248	0.043	5.794	0.000
6pm-10pm	0.300	0.043	6.988	0.000
10pm-3am	0.353	0.064	5.505	0.000
Others_WS	-0.222	0.030	-7.336	0.000
Others_BS	-0.607	0.289	-2.101	0.036
genderf	0.212	0.087	2.440	0.015
age15	-0.033	0.039	-0.860	0.390
COPEc	-0.203	0.074	-2.725	0.006
<i>Random Location (Mean) Effect on WS Variance</i>				
Linear Effect: τ_ℓ	0.894	0.052	17.234	0.000
Quad Effect: τ_q	-0.271	0.031	-8.856	0.000
<i>Random Scale Standard Deviation</i>				
σ_ω	0.621	0.026	24.172	0.000

Table 6: *Variance ratios and 95% confidence intervals for negative affect model*

Variable	Ratio	Lower	Upper
<i>Alpha (BS variance Paramters)</i>			
Intercept	0.849	0.491	1.466
9am-2pm	1.080	0.974	1.196
2pm-6pm	1.107	0.997	1.231
6pm-10pm	1.015	0.912	1.131
10pm-3am	1.058	0.893	1.255
Others_WS	0.739	0.680	0.802
Others_BS	0.588	0.279	1.240
genderf	1.169	0.920	1.485
age15	1.082	0.974	1.201
COPEc	0.435	0.344	0.550
<i>Tau (WS Variance parameters)</i>			
Intercept	1.350	0.873	2.086
9am-2pm	1.191	1.096	1.294
2pm-6pm	1.281	1.178	1.393
6pm-10pm	1.349	1.241	1.468
10pm-3am	1.424	1.255	1.615
Others_WS	0.801	0.755	0.850
Others_BS	0.545	0.310	0.960
genderf	1.236	1.043	1.465
age15	0.967	0.896	1.044
COPEc	0.816	0.706	0.945
<i>Random Location Effect on WS Variance</i>			
Linear	2.445	2.209	2.707
Quadratic	0.763	0.718	0.810
<i>Random Scale Standard Deviation</i>			
Std Dev	1.860	1.769	1.956

2.6.4 Additional analysis

So far, we have only done analysis on the random prompts portion of the study. In the full analysis with the entire data set, we can compare subjects' mood during random prompts and after smoking episodes. Smoking events can be incorporated into the model as an indicator variable, which is then decomposed into BS and WS versions as we did for *Others*. This would allow us to understand the association between tendency to smoke and momentary act of smoking to the mean and variability of mood over and beyond the effect of other covariates which we have analyzed here. Another thing to note is that the data collected did not have detailed time stamps on them so we are not able to fit any kind of models that account for temporal dependence that may exist in the data. With more granular time-stamps, we can fit the mixed location-scale hidden Markov model as discussed in previous section to examine whether there are any discrete latent states whereby adolescent mood responds differently to smoking or being with others.

3 Time-varying effects models (TVEM)

3.1 Motivating study of TVEM

Another way to take advantage of the temporal and contextual information embedded within EMA data sets is to model non-linear, dynamic association between variables unfolding over time, such as modeling the process of relapse as smoking cessation attempts are made under different treatment conditions. Time-Varying Effects Model proposed by Tan et al was developed for this purpose.¹⁶

The motivating study for TVEM is an EMA study from Lanza et al.¹⁷ This study was a randomized, placebo-controlled clinical trial of smoking-cessation therapies. Participants ($N = 1,504$: 58% female; 83% white) smoked at least 10 cigarettes a day for 6 months and were motivated to quit smoking (≥ 8 on a 1-10 scale where 10 is highly motivated to quit). Participants were randomized to monotherapy group, combination therapy group, and placebo group. Each participant had a target quit day (TQD) after receiving treatment or placebo, and over the period of 2 weeks prior and 2 weeks post TQD, four daily EMA measurements (just after waking, prior to bed, and at 2 random times) were collected. The EMA prompts assessed the smoking urge (constructed based on series of items rating the severity of withdrawal symptoms) and negative affect of subject within the last 15 minutes. In addition, baseline nicotine dependence was measured before the study.

The interest is to predict smoking urge with negative affect and baseline nicotine dependence during a quit attempt. Traditionally, the effects of negative affect and nicotine dependence on smoking urge were assumed to be constant over time. However, it may be that the relationship between these predictors and smoking urge is a dynamic process modulated by treatments. For instance, treatments may disentangle the relationship between negative affect and craving over time so that even when subjects are feeling high levels of negative emotions, they will no longer feel the urge to smoke. This would lead to lower craving overall and lower chance of relapse. To model such processes, interaction of polynomial terms of time and predictors could be used. However, when the hypothesized trajectory of changing association is highly nonlinear, such models could be unwieldy to estimate with large numbers of higher order polynomial predictors. In our case, since data were collected over two distinct periods (before and after quit attempt), it is natural to assume an inflection point in the curve at the TQD which is hard to capture with lower order polynomials. To get around this problem, Time-Varying Effects Model (TVEM), based on the varying-coefficient model of Hastie & Tibshirani,¹⁸ models the effect of predictors as non-parametric smooth functions of time using spline based methods.

3.2 Formulation of TVEM

Let y_{it} be subject i 's ($i = 1, 2, \dots, N$) t th ($j = 1, 2, \dots, T_i$) outcome measurement, which we assume to be continuous for simplicity. Let x_{it} be a covariate, which can be time-varying or non-time-varying (in which case x is no longer dependent on t):

$$y_{it} = \beta_0(t_{it}) + \beta_1(t_{it}) \cdot x_{it} + \epsilon_{it} \quad (41)$$

The error ϵ_{it} is assumed to be independent and normally distributed. Both the intercept β_0 and the coefficient β_1 are now functions of time. Therefore, not only are the dependent and independent variables y_{it} and x_{it} changing across time, the coefficients are also changing. At time $t = t_1$, the mean of y_{it} conditional on x_{it} is then $\beta_0(t_1) + \beta_1(t_1)x_{it}$. The goal of TVEM is to estimate the coefficient functions $\beta_0(\cdot)$ and $\beta_1(\cdot)$. Only continuity is assumed and no parametric assumptions are placed on $\beta_0(\cdot)$ and $\beta_1(\cdot)$. The non-parametric approach is more robust to model misspecification, and is suited to capture more complex features such as cyclically. The shortcoming of non-parametric approach is that it requires much more data, which is solved by the intensive sampling done in EMA.

3.3 Estimation of TVEM

3.3.1 Truncated power basis

Various approaches can be taken to estimate the coefficient functions. Here, we take the approach of P-splines (Penalized-splines)¹⁹ with truncated-power basis for ease of computation. Assume data were collected over time interval $[a, b]$. We first divide $[a, b]$ into $K + 1$ intervals. The number of knots K is selected from model selection procedures explained in later sections. Let $\{\tau_1, \tau_2, \dots, \tau_k\}$ be the knots dividing the interval, with $a = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_k < \tau_{k+1} = b$.

Within each of these intervals, we estimate the coefficient functions locally with relatively low-order polynomials formed by linear combinations of the truncated power basis of order q , where q is the degree of the polynomial selected from model selection procedure. The truncated power basis of order q has the form:

$$1, t, t^2, \dots, t^q, (t - \tau_1)_+^q, (t - \tau_2)_+^q, \dots, (t - \tau_k)_+^q \quad (42)$$

where

$$(t - \tau)_+^q = \begin{cases} 0, & \text{if } t \leq \tau \\ (t - \tau)^q, & \text{otherwise} \end{cases} \quad (43)$$

which is called the *truncated power function of order q* with knot τ . Any polynomials formed by this basis is at most of order q , so we are estimating the coefficient functions with piecewise polynomial function of at most order q .

The number of knots K , the placement of knots, and the choice of degree q are all potential tuning parameters. In practice, the degree q is not very important and is usually set relatively low to $q = 2$ or $q = 3$, unless significant lack of model fit suggests otherwise. The number of knots K is selected using information criteria such as AIC or BIC. Placement of knots can either split the interval evenly, or distribute the number of observations falling under each intervals evenly. For simplicity, we take the former approach, which gives us:

$$\tau_r = a + r \frac{b - a}{K + 1}, \quad r = 1, 2, \dots, K \quad (44)$$

Incorporating everything above, the coefficient functions can be approximated as:

$$\beta_0(t) \approx a_0 + a_1 t + a_2 t^2 + \sum_{k=1}^K a_{2+k} (t - \tau_k)_+^2 \quad (45)$$

and

$$\beta_1(t) \approx b_0 + b_1 t + b_2 t^2 + \sum_{k=1}^K b_{2+k} (t - \tau_k)_+^2 \quad (46)$$

with the regression model now written as:

$$y_{it} \approx a_0 + a_1 t_{it} + a_2 t_{it}^2 + \sum_{k=1}^K a_{2+k} (t_{it} - \tau_k)_+^2 + b_0 x_{it} + b_1 t_{it} x_{it} + b_2 t_{it}^2 x_{it} + \sum_{k=1}^K b_{2+k} (t_{it} - \tau_k)_+^2 x_{it} + \epsilon_{it} \quad (47)$$

There are now $2K + 6$ covariates in total: $\{1, t_{it}, t_{it}^2, (t_{it} - \tau_1)_+^2, \dots, (t_{it} - \tau_k)_+^2\}$ and $\{x_{it}, t_{it} x_{it}, t_{it}^2 x_{it}, (t_{it} - \tau_1)_+^2 x_{it}, \dots, (t_{it} - \tau_k)_+^2 x_{it}\}$, with $a_0, a_1, \dots, a_{2+K}, b_0, b_1, \dots, b_{2+K}$ as their coefficients respectively.

We can then write out the estimated coefficient functions as:

$$\hat{\beta}_0(t) = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2 + \sum_{k=1}^K \hat{a}_{2+k} (t - \tau_k)_+^2 \quad (48)$$

and

$$\hat{\beta}_1(t) = \hat{b}_0 + \hat{b}_1 t + \hat{b}_2 t^2 + \sum_{k=1}^K \hat{b}_{2+k} (t - \tau_k)_+^2 \quad (49)$$

3.3.2 Penalizing model complexity

To avoid overfitting, we smooth the estimated coefficient functions using penalty terms. Two approaches have been suggested. Ruppert²⁰ and Wand²¹ have suggested applying penalty terms on the truncated power function terms of the truncated power basis: $\{a_{2+j}, j = 1, 2, \dots, K\}$ and $\{b_{2+j}, j = 1, 2, \dots, K\}$, since these coefficients represent the "jump" in the second order derivatives of the estimated functions at knots relative to the global quadratic polynomial formed by $\{a_0, a_1, a_2, b_0, b_1, b_2\}$. The cost function to be minimized is then:

$$SSE + \lambda_1 \sum_{j=1}^K a_{2+j}^2 + \lambda_2 \sum_{j=1}^K b_{2+j}^2 \quad (50)$$

where

$$SSE = \sum_{i=1}^N \sum_{j=1}^{T_i} \left[y_{it} - \left(a_0 + a_1 t_{it} + a_2 t_{it}^2 + \sum_{k=1}^K a_{2+k} (t_{it} - \tau_k)_+^2 \right) - \left(b_0 + b_1 t_{it} + b_2 t_{it}^2 + \sum_{k=1}^K b_{2+k} (t_{it} - \tau_k)_+^2 \right) \right]^2 \quad (51)$$

If λ_1 and λ_2 are large, then the estimates will be shrunk towards the smooth global second order polynomials $a_0 + a_1 t_{it} + a_2 t_{it}^2$ and $b_0 + b_1 t_{it} + b_2 t_{it}^2$ without any jumps from sub-interval to sub-interval. The tuning parameters λ_1 and λ_2 can be selected using generalized cross-validation and similar algorithms.

Alternatively, when generalized cross-validation algorithms are not readily available in selected computing environment, Wand²¹ has suggested a different approach using mixed-effect model framework. We keep the target of our penalization to the coefficients for the truncated power functions, but we now treat them as random and normally distributed:

$$a_{2+k} \sim N(0, \eta_1), k = 1, 2, \dots, K \quad (52)$$

$$b_{2+k} \sim N(0, \eta_2), k = 1, 2, \dots, K \quad (53)$$

This results in a linear mixed effect model with $a_0, a_1, a_2, b_0, b_1, b_2$ as fixed effects and the rest of the coefficients as random effects. The variance of random effects η_1 and η_2 play similar roles to the tuning parameters λ_1 and λ_2 . If η is 0, then the coefficients do not deviate from their means which is 0, corresponding to infinite regularization. If η is very big, then the coefficients deviate largely from 0, corresponding to minimal regularization.

The joint density of the responses y_{it} and random coefficients $a_{2+k}, k = 1, 2, \dots, K, b_{2+k}, k = 1, 2, \dots, K$ is:

$$P(\mathbf{Y}, a_{2+k}, b_{2+k}, k = 1, 2, \dots, K) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^N \exp \left(\frac{-SSE}{2\sigma^2} \right) \times \left(\frac{1}{\sqrt{2\pi\eta_1}} \right)^K \exp \left(\frac{-1}{2\eta_1} \sum_{k=1}^K a_{2+k}^2 \right) \times \left(\frac{1}{\sqrt{2\pi\eta_2}} \right)^K \exp \left(\frac{-1}{2\eta_2} \sum_{k=1}^K b_{2+k}^2 \right) \quad (54)$$

Where SSE is as defined in equation (45). The log joint density is then:

$$-\frac{N}{2}\log(2\pi\sigma^2) - \frac{K}{2}\log(2\pi\eta_1) - \frac{K}{2}\log(2\pi\eta_2) - \frac{SSE}{2\sigma^2} - \frac{1}{2\eta_1}\sum_{k=1}^K a_{2+k}^2 - \frac{1}{2\eta_2}\sum_{k=1}^K b_{2+k}^2 \quad (55)$$

For the last three terms above, we see that it is equal to equation (44) multiplied by $\frac{1}{2\sigma^2}$ when we denote $\lambda_1 = \frac{\sigma^2}{\eta_1}$ and $\lambda_2 = \frac{\sigma^2}{\eta_2}$. Given this, we see that mixed effect approach is analogous to regularization approach. In fact, Krivobokova & Kauermann²³ has shown that when we use the REML estimates $\hat{\sigma}^2, \hat{\eta}_1, \hat{\eta}_2, \left(\frac{\hat{\sigma}^2}{\hat{\eta}_1}, \frac{\hat{\sigma}^2}{\hat{\eta}_2}\right)$ are comparable to the optimal tuning parameters $(\lambda_1^*, \lambda_2^*)$ selected by generalized cross-validation. Under the mixed-effect model approach, the coefficients for the global polynomial are estimated with maximum likelihood estimates, and the random coefficients for the truncated power functions are estimated by the best-linear unbiased predictor (BLUP) calculated using empirical bayes method same as the one used for MELS.

The advantage of taking the mixed-effect model approach to smooth parameter estimates is that it is easier to implment in varieties of computing environment where GCV algorithms are not readily available. Furthermore, Krivobokova & Kauermann²³ has also shown that this approach is less sensitive to mis-specification of error dependence structure.

3.3.3 Dealing with dependence structure in data

We may add modifications to the model to account for special correlation structures in the data. Firstly, we may assume that the errors are not independent but instead follow an auto-regressive structure, which is a reasonable assumption for EMA data collected over contiguous time points. Let e_{it} denote the autogressive error for observation j of subject i , where ρ is the auto-regression parameter, and $\epsilon_{it} \sim N(0, 1)$:

$$e_{it} = \rho \cdot e_{i(j-1)} + \epsilon_{it} \quad (56)$$

We can reflect this error structure by specifying the corresponding error variance matrix in the joint likelihood which we maximize. Note that in practice, EMA data may be collected over several days, and observation occurring at the beginning of each day may not be related to observation occurring at the end of each day, so simply adding serial auto-correlation to all observations may not make sense. One way to deal with this is to fit a three-level model, whereby observations are nested within days, which are in turn nested within subjects, and consider auto-correlation at different levels. For instance, observations made within the same day could be auto-correlated, while the average value from day to day could also be auto-correlated with a different auto-regression parameter.

Finally, we can add random intercepts and slopes to account for the nested structure of EMA data sets used for TVEM. Let a_i and b_i denote the random intercept and slope for time respectively, each distributed normally with mean 0 and σ_a^2 and σ_b^2 . The model is now:

$$y_{it} = \beta_0(t_{it}) + \beta_1(t_{it})x_{1it} + \beta_p(t_{it})x_{pit} + a_i + b_it_{it} + \epsilon_{it} \quad (57)$$

There are few caveats when incorporating random effects into estimation of TVEM. First, the scale used for time becomes important when we include a random slope for time. This is because the variance of y now depends on time. Assuming that the random effects for slope and intercept are uncorrelated, we have:

$$Var(y_{it}) = \sigma_a^2 + \sigma_b^2 t^2 + \sigma^2 \quad (58)$$

Under this specification, the time point 0 holds the special meaning of time point at which the variance is at minimum. Not only does this affect the interpretation of the coefficients, it also holds implications for model estimation. The random effects are estimated by extrapolating back to time point 0, so if we have used calendar year where the unit is in thousands, the model may have difficulty converging as it tries to extrapolate values back couple of centuries. Therefore, when we include a random slope for time effect in the model, time should be defined as time from some event which is meaningful to the study (such as the start of the study or treatment).

Next, if we are using P-spline with regularization based on the random-effect approach, including more random effects in the form of random slope and intercept may make computations unwildly. In these cases,

we might consider using other methods for regularization, or using B-splines that are more numerically stable.²²

3.3.4 Generalized TVEM

Finally, TVEM can be easily generalized. Let $g(\cdot)$ be the link function. Then the generalized TVEM is:

$$\mathbb{E}(y_{it}) = \mu_{it}, g(\mu_{it}) = \eta_{it} = \beta_0(t_{it})x_{0it} + \beta_1(t_{it})x_{1it} + \dots + \beta_p(t_{it})x_{pit} \quad (59)$$

The coefficient functions are approximated by splines as before which reduces it to regular generalized linear models. Common estimation techniques like maximum likelihood and Generalized Estimation Equation (GEE) can then be applied.

3.4 Data Analysis Example

The data set from the motivating study is not publicly available, so we instead do two simulation studies to demonstrate how TVEM works.

3.4.1 Simulation study on estimating nonlinear coefficient functions

Our first simulation study is designed to examine TVEM's capability in capturing coefficient functions that are non-linear and cyclical over time, which cannot easily be modelled using simple polynomial terms. We generate two data sets, with $N=50$ and $N=200$ respectively. Number of observations per individual (n) is set to 30 for both data sets. Measurement times for each observations are distributed uniformly over $[0, 1]$. Two time-varying covariates x_{1it} and x_{2it} are included, where $x_{1it} \sim \text{Bernouli}(p = 0.5)$ and $x_{2it} \sim N(0, 1)$. The model is specified as:

$$y_{it} = \beta_0(t_{it}) + \beta_1(t_{it})x_{1it} + \beta_2(t_{it})x_{2it} + e_{it} \quad (60)$$

where

$$\beta_0(t) = \exp(2t - 1), \beta_1(t) = 4t^4 - 3t^2 + t, \beta_2(t) = 2\sin^2(2\pi t) \quad (61)$$

the error terms follow an AR(1) structure:

$$e_{it} = \rho e_{i(j-1)} + \epsilon_{it}, \epsilon_{it} \sim N(0, 1) \quad (62)$$

where we set $\rho = 0.5$.

For estimating the model, we use the TVEM SAS macro prepared by Li et al.²² The macro contains options for using P-spline basis with truncated power basis or B-spline. For P-spline, the macro does not provide AIC/BIC statistics as the inclusion of tuning parameters for penalty term complicates the degree of freedom in computing information criteria. While one can still compute AIC/BIC with P-spline by using the "effective degrees of freedom" as defined by Ruppert,²⁰ in practice it is often unnecessary as no matter the number of knots chosen, the penalization term will select a reasonably optimal model by smoothing out the coefficients over the extra knots intervals.

For simplicity, our analysis of the simulated data set will use P-spline method with 10 knots and cubic polynomials ($q = 3$) for the truncated power basis. The estimated coefficient functions along with the true curve and the 95% pointwise confidence bands are shown from Figure 4 to Figure 6. We see that TVEM is able to capture the general shape of the coefficient functions when there are sufficient data. With more data points, the 95% pointwise confidence bands are also narrower.

3.4.2 Simulation study mimicking real data

Our second simulation study uses simulated data set prepared by Li et al²² mimicking a real EMA study to demonstrate how TVEM works in practice. The simulated data set mimics an EMA study following 200 individuals looking to quit smoking. After their target quit day, data was collected over a period of 7 days, with up to 30 random EMA prompts appearing over the period. The EMA prompts collected data about subjects' level of negative affect and smoking urge on a 0 to 10 scale. The EMA prompts also collected measures of subjects' self-efficacy (confidence in their quitting attempts) as well as their current locations. There are two location variables, the first location variable tells whether the subject was at work or at home, and the second location variable tells whether the subject was in a social or fun setting. In addition, the prompts asked subjects whether or not they think their quit attempt will be successful. The interest is to study whether smoking urge has time varying effect on the odds that subjects think their quit attempts will be successful.

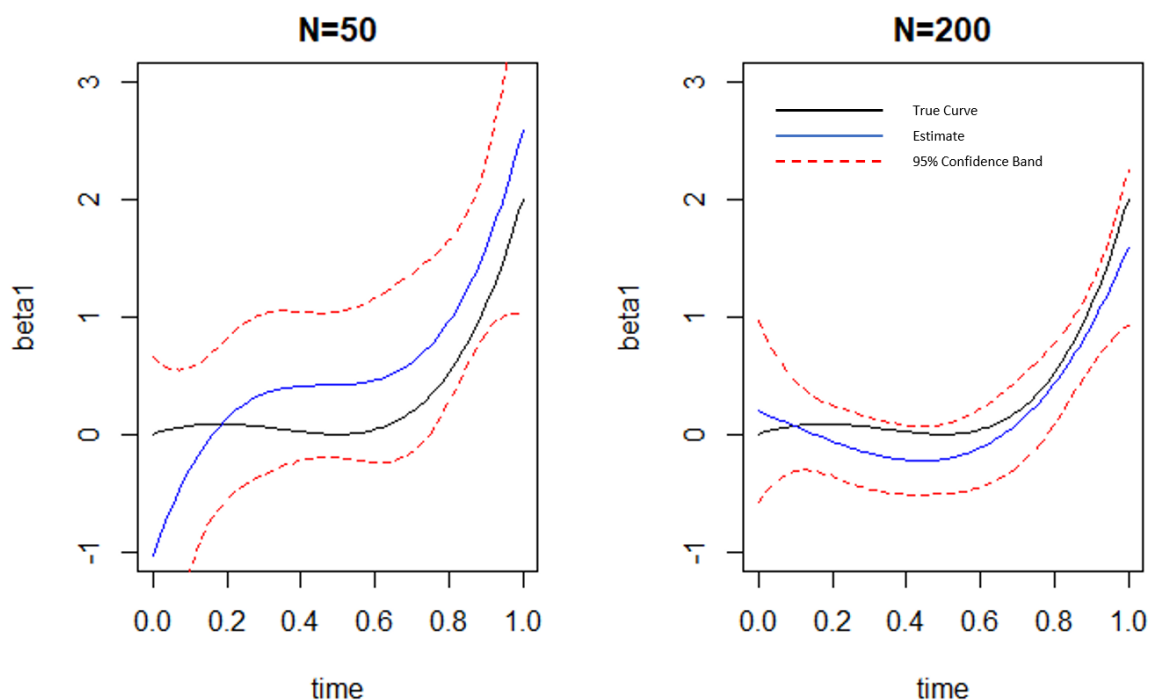


Figure 5: Estimate for $\beta_0(t)$ from first simulation study

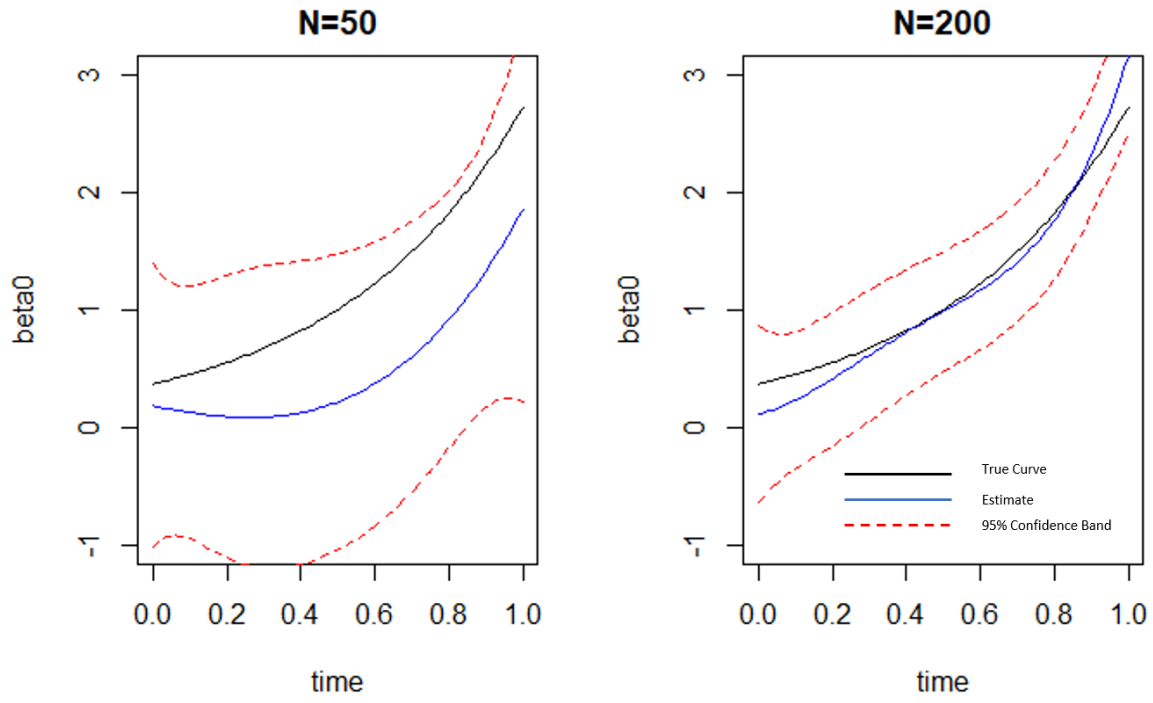


Figure 6: Estimate for $\beta_1(t)$ from first simulation study

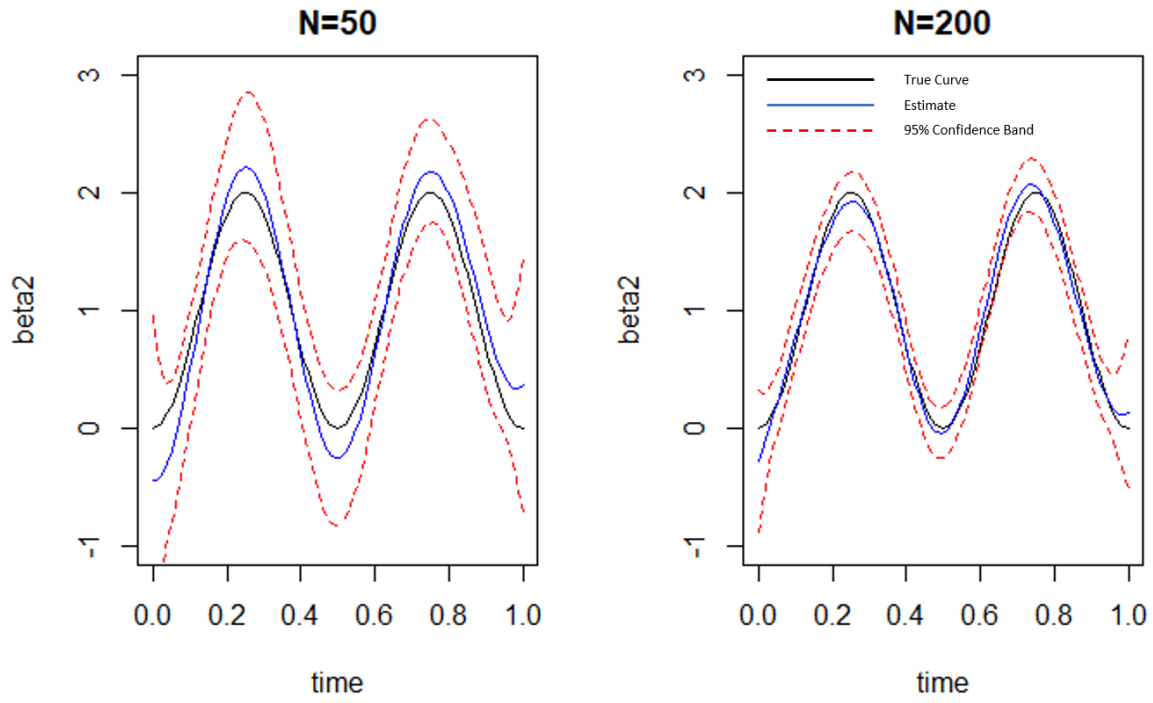


Figure 7: Estimate for $\beta_2(t)$ from first simulation study

We fit the generalized TVEM with logit link, with smoking urge as the time varying effect variable and the location and gender of the subjects as covariates that have time-invariant effects. The model is as specified below:

$$\mathbb{E}(\text{ExpectedSuccess}_{it}) = \mu_{it} \quad (63)$$

Let g be the logit link, we have:

$$g(\mu_{it}) = \eta_{it} = \beta_0(t_{it}) + \beta_1(t_{it})\text{Urge}_{it} + \beta_2\text{location1} + \beta_3\text{location2} + \beta_4\text{gender} \quad (64)$$

Using P-spline with 10 knots, the estimated coefficient functions are plotted in the next page. The exponentiated versions of these curves which represent the odds ratio of expecting success between people differing by one unit of urge. From the curves we see that adjusting for urge, the mean level of confidence declines over time. On the other hand, *Urge* exhibits an increasingly negative effect on confidence as time goes on.

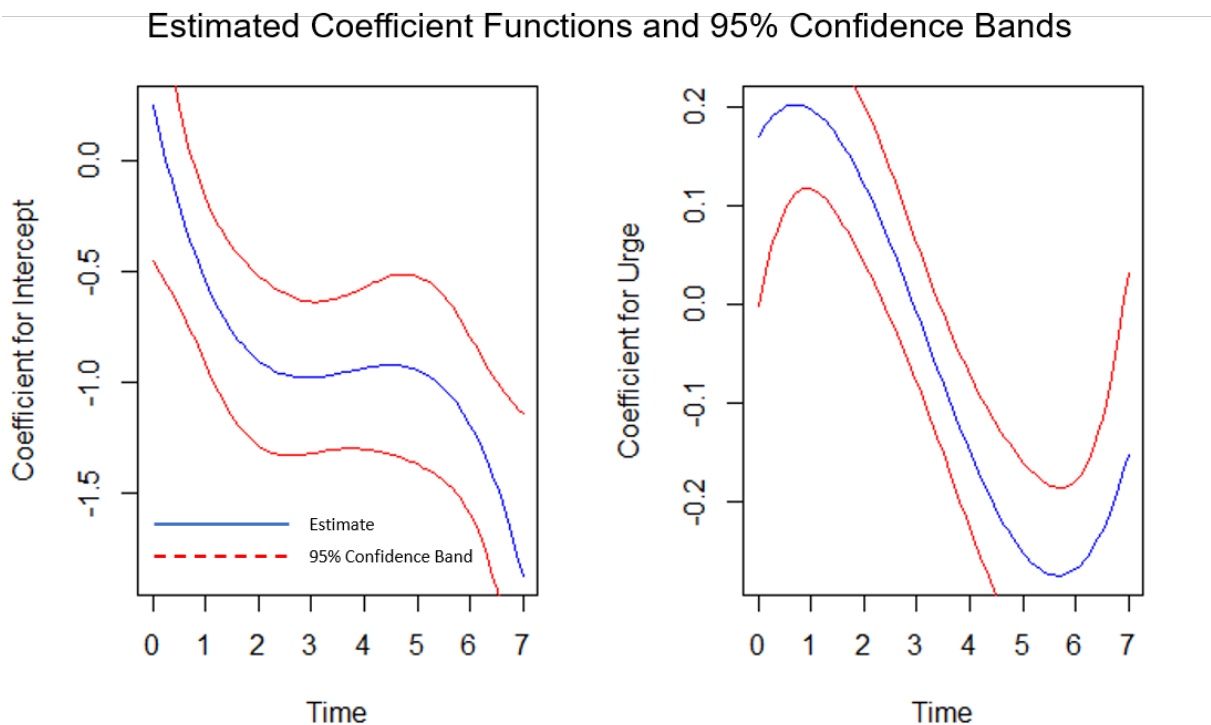


Figure 8: Estimated Coefficient Functions for Intercept and Urge

Estimated Odds Ratio and 95% Confidence Bands

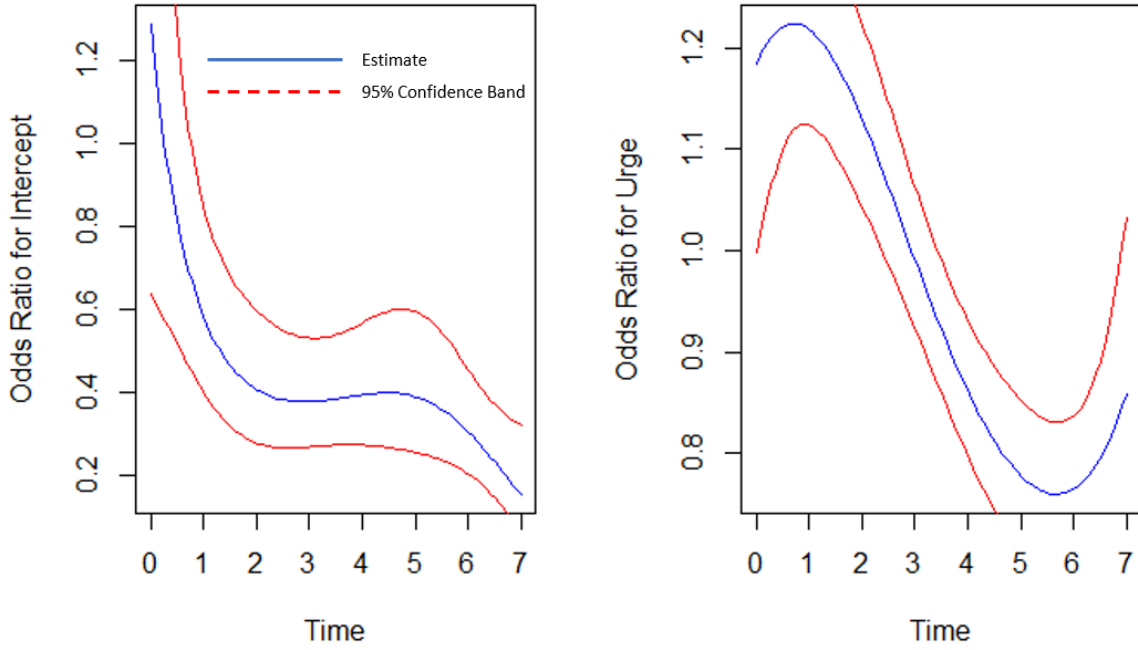


Figure 9: Estimated Odds Ratio for Intercept and Urge

We also report the estimates for time-invariant coefficients below:

Table 7: Estimates for Time-Invariant Coefficients

Effect	Estimate	RobustStdErr	RobustZ	RobustP
Location1	0.00162	0.0682	0.0237	0.981
Location2	0.0196	0.0739	0.265	0.791
Male	0.127	0.0893	1.417	0.156

None of the time-invariant effects are significant at 5% significance level, so the location and gender of the subjects do not have significant effects on the odds of feeling confident about the probability of successfully quitting smoking.

3.4.3 Practical concerns during data analysis

For illustration purposes, in our simulated examples, we have used P-spline with 10 knots as a one-step solution for model selection. We assume that the tuning parameters λ in P-spline method will select an optimally smooth model for us regardless of the number of knots K we picked initially. In practice, P-spline with truncated power basis may encounter numerical instability as the truncated power basis can be highly correlated. Furthermore, P-spline with truncated power basis tend to fit smoother models, so we might want to use other spline basis that are better at catching local patterns. In those cases, we might opt to use the B-spline basis instead by trying out different number of knots and selecting for models manually using AIC/BIC statistics.

Circling back to our original motivating example, the study modelled smoking urge with baseline nicotine dependence (subject-level) and negative affect (occasion-level) as covariates. Separate models were fit for placebo and treatment groups (monotherapy and combination therapy) to compare the shape of coefficient functions over the first 2 weeks post-quit. The study found that the association between negative affect and

craving in treatment groups initially weakened significantly relative to placebo group, but group difference vanished after day 2, whereby negative affect and craving had a positive association. This indicates that the therapies were unable to decouple negative affect with craving. However, the association between baseline nicotine dependence and craving among the combination therapy group was significantly lowered compared to other groups in Day 6-10, confirming the theory that therapies like nicotine replacement helps cessation attempt by reducing the influence of baseline nicotine dependence on withdrawal symptoms such as craving. The study demonstrates TVEM's capacity to reveal the mechanism by which treatments work, and the window of time in which treatments effects are weakened and needs further intervention.

3.5 Extensions of TVEM

An extension we can make with TVEM is to model individually heterogeneous coefficient functions. While TVEM allows for inclusion of random intercept and slope for time, these two only determine the shape of the intercept function $\beta_0(t)$ but not the coefficient functions. To get heterogenous coefficient functions we would have to include random effects for higher order interaction terms, which makes estimation exponentially more cumbersome. At the same time, actually modeling out individual heterogeneity in coefficient functions is too granular to be useful or practical. As we are interested in the general shape of the coefficient functions, subtle differences in individual curves seldom provide meaningful insights. Instead, we may hypothesize that there are distinct latent groups among subjects and that members of the same group share similarly shaped coefficient functions. For instance, in studying the effects of smoking cessation therapy, there may be groups that respond well, do not respond at all, or initially respond well but gradually develops tolerance to the treatment. Li et al²⁴ has proposed MixTVEM (Mixture Time Varying Effects Model) to simultaneously handle heterogeneity and non-linearity of coefficient effects over time.

The model assumes that there are c latent classes among the subjects, each sporting differently shaped coefficient functions. The number of classes c is determined by model selection based on criteria such as AIC and BIC. Conditioning on membership to class c , the model for observation y_{it} is then:

$$\mathbb{E}(y_{it}|c_i = c) = \beta_{0c}(t_{it}) + \beta_{1c}(t_{it})x_{1it} + \dots + \beta_{pc}(t_{it})x_{pit} \quad (65)$$

where the coefficient functions are now sub-scripted by c as they are assumed to vary from class to class.

The latent class membership are modeled using multiple-category logistic regression model. Let $\mathbf{s} = [s_r, r = 1, \dots, R]$ be the vector of R subject-level covariates that have influence on the probability of belonging to each classes, and let $\boldsymbol{\gamma}$ be the vector of coefficients for these covariates. Conditioning on the values of \mathbf{s} , π_{ic} is the probability of subject i to belong to class c , and can be modelled as:

$$\pi_c = \frac{\exp\left(\gamma_{0,c} + \sum_{r=1}^R \gamma_{1,r,c} s_r\right)}{\sum_{c'=1}^C \exp\left(\gamma_{0,c'} + \sum_{r=1}^R \gamma_{1,r,c'} s_r\right)} \quad (66)$$

The coefficients \mathbf{s} in the multiple-category logistic regression model are estimated using EM algorithm just as in other mixture models. The clustered structure of data is taken into consideration, and individuals are assumed to belong to only one single class without switching classes across observations. In the initial E-step, random initial probabilities of individuals belonging to each class are generated. These probabilities are used as weights to estimate the multiple-category logistic regression model in the M-step, and the algorithm iterates itself until convergence. A key tuning parameter for the model is the number of latent classes. There are certain philosophical ambiguities in the conceptualization of these latent classes. Under certain contexts, there could be a number of clinically defined categories of subjects, and we could use that number directly. In other cases, there are no well-defined boundaries or rules for classifying subjects, and we let the model decide. In these cases, we can use information criteria such as AIC and BIC to determine the appropriate number of classes. Alternatively, we can also use the Weighted Version of Residual Sum of Squared Errors (WRSS):

$$\text{WRSS} = \sum_{i=1}^N \sum_{j=1}^{T_i} \sum_{c=1}^{n_{class}} \omega_{ic} \left(\sum_{j=1}^{T_i} (y_{it} - \hat{y}_{it|c})^2 \right) \quad (67)$$

where N is the number of individuals, T_i is the number of observations for the i th individual, $n_{classes}$ is the number of latent classes, ω_{ic} is the posterior probability of class membership for individual i and class c , and $\hat{y}_{it|c}$ is the predicted value for y_{it} conditioning on membership to class c . In practice, an "elbow plot" of WRSS against number of classes is drawn, and the point in which a decrease in model size lead to significant worsening of model performance is chosen.

4 Discussions

As discussed in the paper, both MELS and TVEM are developed to unlock the contextual and temporal information embedded in EMA data. The intensively repeated and time-stamped sampling of individual data yield enough information for modeling variances at both between-subject and within-subject level in MELS, as well as for modeling covariate effects as nonlinear functions of time in TVEM.

In the case of MELS and MixTVEM, computation can be burdensome due to the inclusion of multiple random effects. For MELS, it is also difficult to estimate a model with complex auto-correlation structure in level-1 error terms. When strong auto-correlation in error terms are ignored, estimates and inferences done on variance sub-models in MELS could be invalid. Mixed location-scale HMM offers an alternative framework to account for serial dependence of observations for MELS, but its assumptions are not applicable to all psychological and clinical processes.

In the future, EMA data will be increasingly integrated with mHealth analytic platforms, where integration with other data sources such as GPS location data and continuously collected physiological data could extend the range of questions EMA data can answer. Ultimately, the goal is to analyze EMA data sequentially as they are being generated to develop individually tailored, adaptive interventions that can be delivered in real-time via personal devices.²⁵ For instance, in "DietAlert" app developed by Goldstein and colleagues,²⁶ a predictive model for dietary lapses after weight control diet was developed to send tailored intervention (e.g., a text message) to subjects based on their EMA recordings to the app. EMA recordings of past lapse instances were used as training data for the model, whereby contextual information, individual baseline traits, and historical trajectories were all used as predictors for predicting whether lapse would occur. Traditional models such as MELS and TVEM work to explain rather than to predict. However, insights gained from these models could in turn be used to build predictive models based on the same EMA data set. For instance, in mixed location-scale hidden Markov model, the dominant latent state of individuals could inform the development of individually tailored interventions. In TVEM, kinks in the shape of estimated coefficient effect functions over time could inform the time points at which intervention is most needed. Combining with other data sources such as physiological data and EHR data, we could further zero in on the optimal time of intervention. With all that being said, looking for ways to effectively integrate analysis from traditional EMA models with the influx of new data streams coming into mHealth platforms could be one future direction of research in ways to extract more insights from EMA data.

A Simulation codes for MELS

The code for generating the data used in the simulation study of MELS is as below:

```
# Function to generate EMA - like data:
DataGen <- function(nsubj, nobs,
                    beta = c(1,1,1),
                    tau = c(0.5,0.6,-0.4),
                    sd_subj_loc = 1,
                    sd_subj_scale = 0.8,
                    tau_ell = 0.3){
  beta0 = beta[1]; beta1 = beta[2]; beta2 = beta[3]
  tau0 = tau[1]; tau1 = tau[2]; tau2 = tau[3]

  # Generate subject wave and time indicators
  subject =rep(seq(1:nsubj),each = nobs)

  # Generate subject, wave and obs level covariates (observed).
  # subject-level covariate value for each subject
  x_subj_orig <- rbinom(nsubj,1,0.5)
  # Prepare subject-level variables to design matrix format
  x_subj <- rep(x_subj_orig,each=nobs)
  x_obs <- rnorm(N)

  # Generate random subject effects for each subject
  theta1 = rep(rnorm(nsubj), each = nobs)
  theta2 = rep(rnorm(nsubj), each = nobs)

  # Generate the mean
  y_mean <- beta0 + beta1 * x_subj + beta2 * x_obs + sd_subj_loc * theta1
  # Generate the sd
  y_sd <- sqrt(exp(tau0 + tau1 * x_subj + tau2 * x_obs
                  + tau_ell * theta1 + sd_subj_scale *theta2))

  # Generate response
  y <- rnorm(N, mean = y_mean, sd = y_sd)
  # Return a data frame
  df_LSME <- list(data = data.frame(subject = subject,
                                    x1 = x_subj, x2 = x_obs,y = y),
                  x_subj_orig = x_subj_orig)

  return(df_LSME)
}
```

Code for function that prepares inputs to be used by Stan and produce MCMC fits:

```
run_MCMC <- function(data, model, iter = 1000, warmup = 500, chains = 4){

  model_names = c("MEL.stan","MELS.stan")
  model_name = model_names[model] # flag for which model to use

  # Prepare data for Stan
  df.LSME <- data$data
  # subject-level variable in original format
  x_subj_orig <- data$x_subj_orig
  nsubj <- length(unique(df.LSME$subject))
  nobs <- 25
  N <- nrow(df.LSME)
  x1 <- df.LSME$x1 # subject level variable
  x2 <- df.LSME$x2 # occasion level variable
  subject <- as.integer(df.LSME$subject) # subject ids
  # Design matrix:
  X_mean <- unname(model.matrix(~ 1 + x1 + x2)) # design matrix for mean
  X_var <- unname(model.matrix(~ 1 + x1 + x2)) # design matrix for variance
  y <- df.LSME$y

  # Combine to list format for Stan
  StanDat <- list (N = N, nsubj = nsubj, subject = subject,
                  X_mean = X_mean, X_var = X_var, y = y)

  library(lme4)
  lme_fit <- lmer(y ~ x1 + x2 + (1|subject))
  sum <- summary(lme_fit)
  beta_init = sum$coefficients[, "Estimate"]
  sigma_subj_loc_init = sqrt(diag(VarCorr(lme_fit)$subject))
  tau_init = c(log(sum$sigma), 0, 0)

  # model 1 <=> model without scale effect
  if(model == 1){
    MLS_init <- function(){
      list(beta = beta_init,
           sigma_subj_loc = sigma_subj_loc_init,
           tau = tau_init,
           z_subj_loc = rnorm(nsubj))
    }
  }
  # model 2 <=> model with scale effect
  }else if(model == 2){
    MLS_init <- function(){
      list(beta = beta_init,
           sigma_subj_loc = sigma_subj_loc_init,
           tau = tau_init,
           tau_ell = 0.1,
           sigma_subj_scale = 0.1,
           z_subj_loc = rnorm(nsubj),
           z_subj_scale = rnorm(nsubj))
    }
  }
}
```



```

MLS.fit <- stan(data = StanDat,
               file = model_name,
               init = MLS_init,
               iter = iter,
               warmup = warmup,
               chains = chains)

return(MLS.fit)
}

```

The Stan code for estimating a model without scale effect:

```

data {
  int<lower=1> N; //number of data points
  int<lower=1> nsubj; //number of subjects
  int<lower=1, upper=nsubj> subject[N]; //indicator for subjects
  row_vector[3] X_mean[N]; //design matrix for fixed effect
  row_vector[3] X_var[N]; //design matrix for fixed effect
  real y[N]; //outcome
}

parameters {
  vector[3] beta; //fixed effect for mean
  real<lower=0> sigma_subj_loc; // location effect sd
  vector[3] tau; // fixed effect for subject level location RE variance
  real z_subj_loc[nsubj]; // standardized random location effect
}

model {
  //REs
  for(i in 1:nsubj){
    z_subj_loc[i] ~ normal(0,1);
  }

  // likelihood
  for (i in 1 : N) {
    y[i] ~ normal(X_mean[i] * beta + sigma_subj_loc * z_subj_loc[subject[i]],
                  sqrt(exp(X_var[i] * tau)));
  }
}

```

The Stan code for estimating model with scale effect:

```

data {
  int<lower=1> N; //number of data points
  int<lower=1> nsubj; //number of subjects
  int<lower=1, upper=nsubj> subject[N]; //indicator for subjects
  row_vector[3] X_mean[N]; //design matrix for fixed effect
  row_vector[3] X_var[N]; //design matrix for fixed effect
  real y[N]; //outcome
}

```

```

parameters {
  vector[3] beta; //fixed effect for mean
  real<lower=0> sigma_subj_loc; //fixed effect for variance
  vector[3] tau; // fixed effect for subject level location RE variance
  real tau_ell; // linear effect of location effect on WS-variance
  real<lower=0> sigma_subj_scale; // standard deviation of scale effect
  real z_subj_loc[nsubj]; // standardized random location effect
  real z_subj_scale[nsubj]; // standardized random scale effect
}

model {
  //REs
  for(i in 1:nsubj){
    z_subj_loc[i] ~ normal(0,1);
    z_subj_scale[i] ~ normal(0,1);
  }

  // likelihood
  for (i in 1 : N) {
    y[i] ~ normal(X_mean[i] * beta + sigma_subj_loc * z_subj_loc[subject[i]],
      sqrt(exp(X_var[i] * tau + tau_ell * z_subj_loc[subject[i]] +
        sigma_subj_scale * z_subj_scale[subject[i]])));
  }
}

```

B Simulation codes for mixed location-scale HMM

The code for generating the data used in the simulation study of mixed location-scale HMM is as below:

```
# Set Simulation Parameters
niter_t <- 2000
nwarmup_t <- 1000
nchains_t <- 4

## Set simulation parameter
nsubj_t <- 100 # Num timeseries.
nt_t <- 10 # Length of timeseries.
p0_t <- 0.5 # Initial prob state=1
tp_t <- matrix(c(0.9,0.1,0.3,0.7),nrow = 2, byrow = TRUE) # Transition probabilities

mu.1 <- 1
mu.2 <- 2
mu_t <- c(mu.1, mu.2) # Emission location parameter

sd.1 <- 2
sd.2 <- 1
sig_e_t <- c(sd.1, sd.2) # Emission scale parameter

cor_ls <- -0.3
var_ls <- c(1.0, 0.3) # Random subject effect variance
cov_ls_t <- matrix(c(var_ls[1], sqrt(var_ls[1]) * sqrt(var_ls[2]) * cor_ls,
                    sqrt(var_ls[1]) * sqrt(var_ls[2]) * cor_ls, var_ls[2]),
                  nrow = 2, byrow = TRUE)

## Data generation
DataGen_HMM <- function(nsubj = nsubj_t, nt = nt_t, p0 = p0_t, tp = tp_t,
                        mu = mu_t, sig_e = sig_e_t, cov_ls = cov_ls_t){
  # Initialise state and data structures.
  Z <- rep(NA, nsubj * nt)
  dim(Z) <- c(nsubj, nt)
  # Fill first column.
  Z[, 1] <- sample(c(1, 2),
                  replace = TRUE,
                  prob = c(p0, 1 - p0),
                  size = nsubj)

  # table(Z[, 1])
  # Fill subsequent columns.
  for (t in 2 : nt){
    for(i in 1 : nsubj){
      Z[i, t] <- sample(c(1, 2),
                      size = 1,
                      prob = tp[Z[i, t - 1], ])
    }
  }
  # Generate random effect and data.
  re_ls <- mvrnorm(nsubj, mu = c(0, 0), Sigma = cov_ls)
  re_loc <- re_ls[, 1]
  re_scale <- re_ls[, 2]
  y <- matrix(NA, nrow = nsubj, ncol = nt)
```

```

for (i in 1 : nsubj){
  for(t in 1 : nt){
    y[i, t] <- rnorm(1, mu[Z[i,t]] + re_loc[i], sig_e[Z[i,t]] * sqrt(exp(re_scale[i])))
  }
}

# Output generated data
sim.df <- data.frame(subject = rep(seq_len(nsubj), each = nt),
                     time = rep(seq_len(nt), nsubj),
                     y = as.vector(t(y)))

# Return data set
return(list(re_loc = re_loc,
           re_scale = re_scale,
           y = y,
           sim.df = sim.df))
}

```

Code for function that prepares inputs to be used by Stan and produce MCMC fits:

```

run_MCMC_HMM = function(data, model_num, iter = niter_t,
                        warmup = nwarmup_t, chains = nchains_t){
  model_names = c("hmm.stan", "hmm_l.stan", "hmm_ls.stan")
  model_file = model_names[model_num]
  data.list = data
  StanDat <- list(K = 2,
                 nsubj = length(unique(data.list[[4]]$subject)),
                 nt = length(unique(data.list[[4]]$time)),
                 y = data.list[[3]])

  ## Get initial value
  InitFit <- lmer(y ~ (1 | subject), data = data.list[[4]])
  sum <- summary(InitFit)

  if(model_num==3){
    StanInit <- function(){list(p0 = 0.5, tp = c(0.8, 0.8),
                                mu = c(sum$coefficients[1, "Estimate"] - 0.1,
                                      sum$coefficients[1, "Estimate"]) + 0.1,
                                sig_e = c(sum$sigma + 0.5, sum$sigma - 0.5),
                                sig_ls = c(sqrt(diag(VarCorr(InitFit)$subject)), 0.01),
                                chol_ls = chol(matrix(c(1, -0.3, -0.3, 1),
                                                       nrow = 2, byrow = T)),
                                theta = array(rnorm(2*nsubj_t), dim = c(nsubj_t, 2)))}
  }else if(model_num==1){
    StanInit <-function(){ list(p0 = 0.5, tp = c(0.8, 0.8),
                                mu = c(sum$coefficients[1, "Estimate"] - 0.1,
                                      sum$coefficients[1, "Estimate"]) + 0.1,
                                sig_e = c(sum$sigma + 0.5, sum$sigma - 0.5))}
  }else if(model_num==2){
    StanInit <-function(){ list(p0 = 0.5, tp = c(0.8, 0.8),
                                mu = c(sum$coefficients[1, "Estimate"] - 0.1,
                                      sum$coefficients[1, "Estimate"]) + 0.1,
                                sig_e = c(sum$sigma + 0.5, sum$sigma - 0.5))}
  }
}

```

```

        sig_e = c(sum$sigma + 0.5, sum$sigma - 0.5),
        sig_loc = sqrt(diag(VarCorr(InitFit)$subject)),
        theta = rnorm(nsubj_t))}
}

# Get model fit:
HmmFit <- stan(data = StanDat,
              file = model_file,
              init = StanInit,
              iter = iter,
              warmup = warmup,
              chains = chains,
              cores = 10)

return(HmmFit)
}

```

The Stan code for estimating mixed location-scale HMM

```

data {
  int<lower=1> K; // number of groups
  int<lower=1> nsubj; // number of data points
  int<lower=1> nt; // length of timeseries
  matrix[nsubj,nt] y; // observations
}

parameters {
  real<lower=0,upper=1> p0 ; // initial prob grp 1
  vector<lower=0,upper=1>[K] tp ; // transition probs of staying in group
  ordered[K] mu; // location parameter of mixture components
  vector[K] sig_e; // scale parameter of mixture componets
  vector<lower=0>[2] sig_ls; // SD (scales) of mixture components
  cholesky_factor_corr[2] chol_ls; // Cholesky factors of random effects covariance matrix
  vector[2] theta[nsubj]; // standardized REs (location and scale)
}

transformed parameters {
  matrix<lower=0,upper=1>[nsubj,nt] pred; // one-step filter prediction of group
  vector[2] nu[nsubj]; // RE array (location and scale)
  real nu_loc[nsubj]; // RE location
  real nu_scale[nsubj]; // RE scale
  vector[2] Sig_ls; // Variance of random effects
  real cor_ls; // Correlation of random effects
  {
    matrix[nsubj,nt] F; //filtered belief states
    real like1;
    real like2;
    real p1;
    real p2;

    Sig_ls = sqrt(diagonal(crossprod(diag_pre_multiply(sig_ls, chol_ls))));
    cor_ls = crossprod(chol_ls)[1,2];
  }
}

```

```

//Forwards algorithm
for (n in 1:nsubj){
  F[n, 1] = p0;
  pred[n, 1] = F[n, 1];
  nu[n] = (diag_pre_multiply(sig_ls, chol_ls)) * theta[n];
  nu_loc[n] = nu[n][1];
  nu_scale[n] = nu[n][2];
}
for (t in 1:nt){
  for (n in 1:nsubj) {
    //update prior using data
    like1 = exp(normal_lpdf(y[n, t] | mu[1] + nu_loc[n],
                           sig_e[1] * sqrt(exp(nu_scale[n]))));
    // local evidence for class 1
    like2 = exp(normal_lpdf(y[n, t] | mu[2] + nu_loc[n],
                           sig_e[2] * sqrt(exp(nu_scale[n]))));
    // local evidence for class 2
    p1 = F[n, t] * like1; // joint for class 1
    p2 = (1 - F[n, t]) * like2; // joint for class 2 (used for normalizing constant)
    F[n, t] = p1 / (p1 + p2); // update filtered belief state

    //predict forward one timestep
    if (t != nt) {
      p1 = F[n, t] * tp[1] + (1 - F[n, t]) * (1 - tp[2]);
      // one-step ahead predictive prob for class 1
      p2 = F[n, t] * (1 - tp[1]) + (1 - F[n, t]) * tp[2];
      // one-step ahead predictive prob for class 2
      F[n, t+1] = p1 / (p1 + p2); // prep for computing update next cycle
      pred[n,t+1] = F[n,t+1]; // prep for likelihood computation
    }
  }
}

}

model {
  // declare temp for log component densities
  real ps;
  // Priors
  p0 ~ uniform(0, 1);
  tp ~ uniform(0, 1);
  mu ~ normal(0, 100);
  chol_ls ~ lkj_corr_cholesky(1);
  for(n in 1 : nsubj){
    theta[n] ~ normal(0, 1);
  }
  // Likelihood
  for (n in 1 : nsubj){
    for (t in 1:nt) {
      ps = pred[n, t] * exp(normal_lpdf(y[n, t] | mu[1] + nu_loc[n],
                                         sig_e[1] * sqrt(exp(nu_scale[n])))) +

```

```

        (1 - pred[n, t]) * exp(normal_lpdf(y[n, t] | mu[2] + nu_loc[n],
                                          sig_e[2] * sqrt(exp(nu_scale[n]))));
    target += log(ps);
  }
}

```

References

- ¹ Shiffman, S., Stone, A., and Hufford, M. (2008), "Ecological momentary assessment", *Annual Review of Clinical Psychology*, 4: 1-32. doi:10.1146/annurev.clinpsy.3.022806.091415.
- ² Bradburn, N., Rips, L., Shevell, S. (1987). Answering autobiographical questions: the impact of memory and inference on surveys. *Science*, 236:157-61.
- ³ Kubiak, T, Smyth, J. (2019), "Connecting Domains- Ecological Momentary Assessment in a Mobile Sensing Framework". In H. Baumeister and C. Montag(Eds.), *Digital Phenotyping and Mobile Sensing* (pp.201-20). Springer Nature, Switzerland, AG.
- ⁴ Mermelstein, R., Hedeker, D., Flay, B., Shiffman, S. (2002). Situational versus intra-individual contributions to adolescents' subjective mood experience of smoking. Savannah, GA: Annual Meeting for the Society for Research on Nicotine and Tobacco.
- ⁵ Hedeker, D., Mermelstein, R.J. (2020). Modeling Variation in Intensive Longitudinal Data. In A.A. O'Connell, D.B.McCoach, & B.Bell (Eds.), *Multilevel Modeling Methods with Introductory and Advanced Applications*. Information Age Publishing.
- ⁶ Hedeker, D., Gibbons, RD. (2006). *Longitudinal Data Analysis*. New York, NY: Wiley.
- ⁷ Hedeker, D., Nordgren, R. (2013). Mixregls: A program for mixed-effects location scale analysis. *Journal of Statistical Software*, 52(12):1-38.
- ⁸ Ma, Q. & Hedeker, D. (2020). Modeling of between- and within-subject variances using mixed effects location scale (MELS) models. SAS 2020 Global Forum Proceedings, Paper 4181-2020.
- ⁹ Gelman, A., Carlin, JB., Stern, HS., Dunson, DB., Vehtari, A., Rubin, DB. *Bayesian Data Analysis*. Florida: CRC Press; 2013.
- ¹⁰ Dzubur, E., Ponnada, A., Nordgren, R., Yang, C.-H., Intille, S., Dunton, G., Hedeker, D. (2020). MixWILD: A program for examining the effects of variance and slope of time-varying variables in intensive longitudinal data. *Behavior Research Methods*, 52(4):1403-1427.
- ¹¹ Chi, E.M., Reinsel, G.G: Models for longitudinal data with random effects and ar(1) errors. *J. Am. Stat. Assoc.* 84(406), 452-459 (1989)
- ¹² Ma, Q., Mermelstein, R., Hedeker, D. (2020). A three-level mixed model to account for the correlation at both the between-day and the within-day level for ecological momentary assessments. *Services and Outcomes Research Methodology*. doi: 10.1007/s10742-020-00220-w.
- ¹³ Lin, X., Mermelstein, R., Hedeker, D. (2020). Mixed location scale hidden Markov model for the analysis of intensive longitudinal data. *Health Services and Outcomes Research Methodology*. doi: 10.1007/s10742-020-00217-5.
- ¹⁴ Liu, Y., et al. (2015). Efficient learning of continuous-time hidden Markov models for disease progression. *Adv. Neural Inf. Process. Syst.* 28, 3599-3607.
- ¹⁵ Watson, D., Wiese, D., Vaidya, J., Tellegen, A. (1999) The two general activation systems of affect: structural findings, evolutionary considerations, and psychobiological evidence. *J Pers Soc Psychol.* 76:820-38.
- ¹⁶ Tan, X., Shiyko, M., Li, R., Li, Y., & Dierker, L. (2012). A time-varying effect model for intensive longitudinal data. *Psychological Methods*, 17(1): 61-77. doi:10.1037/a0025814 PMID: PMC3288551.
- ¹⁷ Lanza, ST., Vasilenko, SA., Liu, X., Li, R., & Piper, M. (2014). Advancing the understanding of craving during smoking cessation attempts: A demonstration of the time-varying effect model. *Nicotine and Tobacco Research*, 16:S127-134.

- ¹⁸ Hastie, T., Tibshirani, R. (1993). Varying-coefficient models. *Journal of the ROyal Statistical Society, Series B.* 55(4): 757-779.
- ¹⁹ Rupper D., Wand, MP., Carrol, RJ. (2003). Semiparametric Regression. New York: Cambridge University Press.
- ²⁰ Ruppert D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics.* 11(4): 735-757.
- ²¹ Wand MP. (2003) Smoothing and mixed models. *Computational Statistics.* 18:223-249.
- ²² Li, R., Dziak, J.D., Tan, X., Huang, L., Wagner, A.T., & Yang, J. (2017). *TVEM (time-varying effect modeling) SAS macro user's guide* (Version 3.1.1). University Park: The Methodology Center, Penn State. Retrieved from <http://methodology.psu.edu>
- ²³ Krivobokova, T., Kauerman, G. (2007). A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association.* 93: 1328-1337.
- ²⁴ Li, R., Dziak, J.J., Tan, X., Shiffman, S., Shiyko, M.P. (2015). Modeling Intensive Longitudinal Data With Mixtures of Nonparametric Trajectories and Time-Varying Effects. *Psychological Methods.* 20(4): 444-469.
- ²⁵ Jerome, A., Fiero, P.L., & Behar, A. (2000). Computerized scheduled gradual reduction from smokeless tobacco cessation: Development and preliminary evaluation of a self-help program. *Computers in Human Behavior.* 16:493-505. doi: [http://dx.doi.org/10.1016/S0747-5632\(00\)00024-8](http://dx.doi.org/10.1016/S0747-5632(00)00024-8)
- ²⁶ Goldstein, S., et al. (2017). Return of the JITAI: applying just-in-time adaptive intervention framework to the development of m-health solutions for addictive behaviors. *Int J Behav Med.* 24(5):673-82.