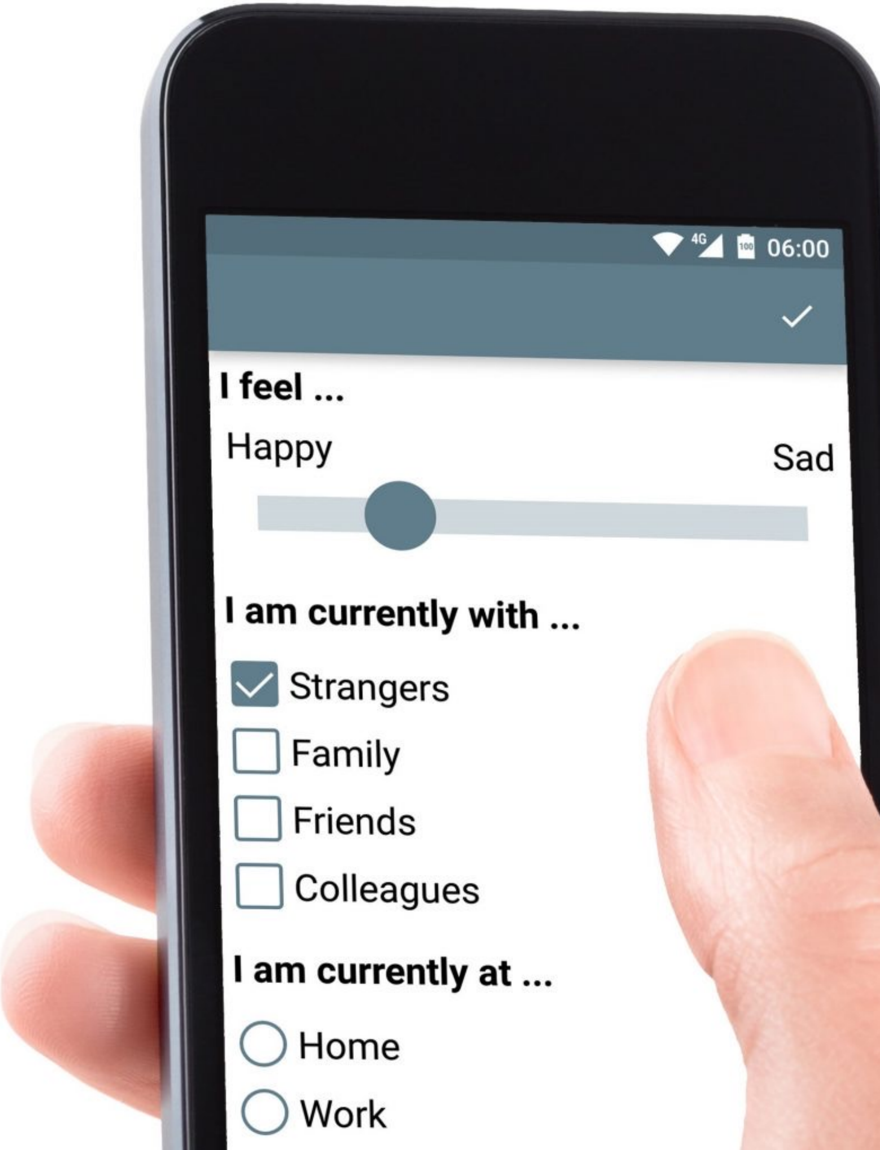
Modeling Ecological Momentary Assessment Data Using Mixed-Effects Location-Scale Model and Time Varying Effects Model

Adam Zhang

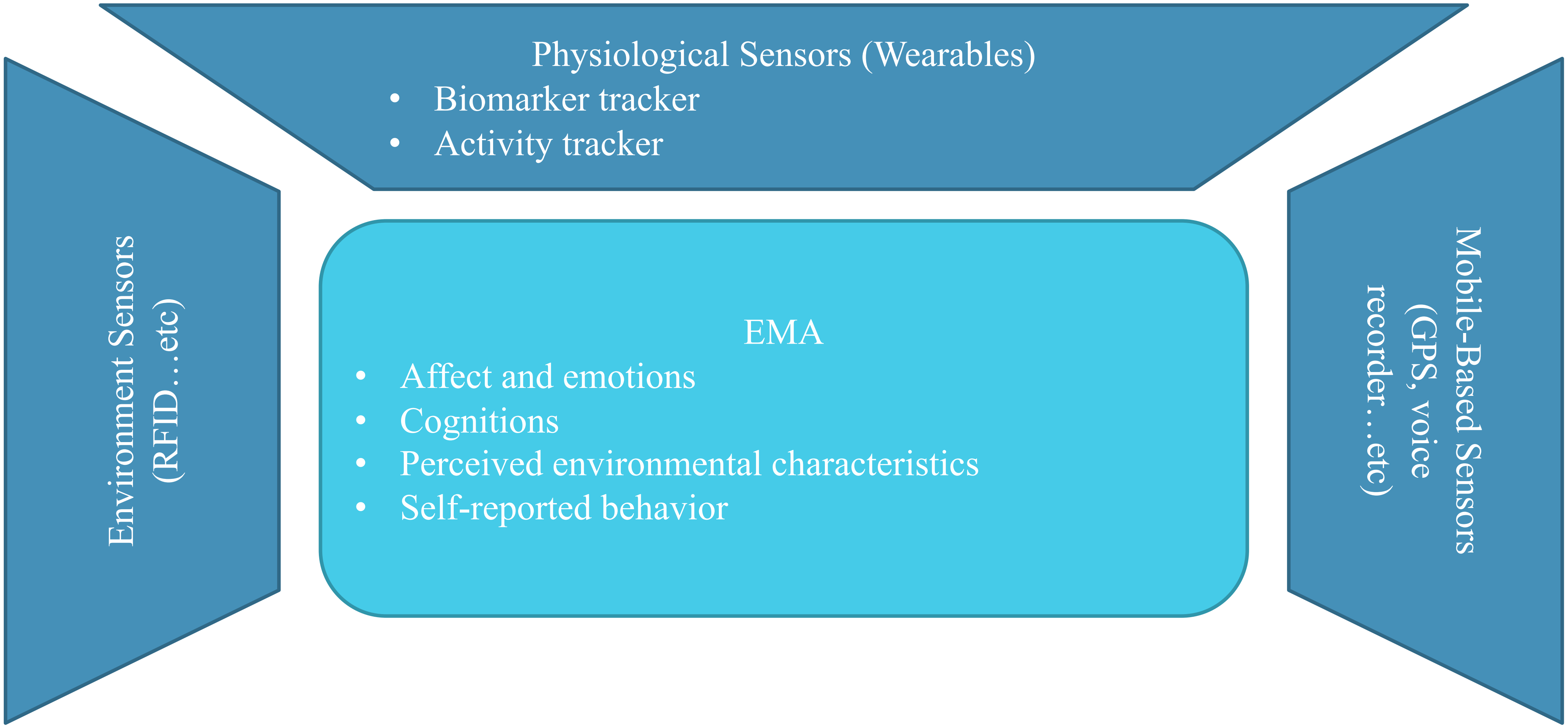
What is Ecological Momentary Assessment?

- Using real-time data capture techniques such as smartphone prompts to repeatedly sample subjects' behaviors, thoughts, and states in their natural environment
  - Less recall bias
  - Collects contextual and momentary information needed to understand how subject behaviors and experiences unfold and vary over time in real world
  - Wide applications in psychology, sociology, and public health sciences
- As smartphone becomes more prevalent, number of studies using EMA has increased rapidly as well



EMA as part of mHealth platform

EMA is an integral component of mobile sensing framework in mHealth platform



Use of EMA Data

Modeling

- Infer an ecological explanation of the reasons for maladaptive behaviors (smoking, drug use, binge eating…etc) based on reliable patterns of change (e.g., affective states from prior to and following a behavior)

Modeling Variability:

- Jointly model covariate effects on both the mean and variance of response variable (usually mood)

Non-linear association between variables

- Model changing nature of associations between key variables throughout some clinical process of interest such as smoking cessation attempt

Predicting

- Classifying subject’s emotional / behavioral states
- Predicting imminent relapse based on EMA entry collected

Intervention

- Just-In-Time Adaptive Intervention to deliver appropriate support at the right timing / place according to EMA entry
- Update decision rules dynamically

Modeling variance in EMA Data

A motivating study

EMA portion of a longitudinal natural history study of adolescent smoking (Mermelstein et al., 2002)

- **Participants**: 510 adolescents who were in 8th or 10th grade at the beginning of the study

  Were either non-smokers who indicated interest in smoking, or light smokers
- **Data collection:** Subjects carried hand-held computers over 7 days
  - Asked to respond to interview prompts appearing randomly (avg. 7 prompts per day)
  - Also asked to initiate the data collection interview whenever smoking
  - Interview prompts:
    - Range of subjective mood evaluation items (e.g., I felt happy, I felt relaxed, I felt cheerful…etc), rated from 1 to 10
      - Positive / negative affect measured by averaging scores on mood items
      - Circumstances of smoking: with other people or alone

Questions of Interest:

1. Comparing data across random prompt portion and smoking-initiated portion, what is the effect of smoking on the mean and variability of either negative and positive affect?
2. How do we account for individual heterogeneity in both mean (location) and variability (scale) to carry out correct inference?
3. How do we account for potential serial correlation in the data?

Formulation of mixed effects location-scale model (1/2)

Mean sub-model:

Let y_{it} denote the measurement of response variable such as positive /negative affect level for subject i $(i = 1,2, \dots, N)$ on occasion t $(t = 1,2, \dots, T_i)$

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

$\boldsymbol{x}_{\boldsymbol{it}}$ = p-dimensional vector of regressors

$\boldsymbol{\beta}$ = p-dimensional vector of regression coefficients

$\nu_i \sim N(0, \sigma_\nu^2)$: Random location effect; Between-subject (BS) variance - how heterogeneous subjects are in mean?

$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$: Error term, assumed independent across subjects and occasions for now ; Within-subject (WS) variance – how erratic are the data within subjects?

Variance sub-models in log-linear representation:

$$\sigma^2_{\epsilon_{it}} = \exp(\boldsymbol{\omega_{it}}'\boldsymbol{\tau}) \qquad \sigma^2_\nu = \exp(\alpha_0)$$

$\boldsymbol{w}_{\boldsymbol{it}}$ = q-dimensional vector of regressors for WS-variance sub-model

$\boldsymbol{\tau}$ = q-dimensional vector of regression coefficients for WS-variance sub-model

α_0 = Intercept of BS-variance sub-model

Formulation of mixed effects location-scale model (2/2)

Adding scale random-effect to WS-variance sub-model

$$\sigma^2_{\epsilon_{it}} = \exp(\boldsymbol{\omega}'_{\boldsymbol{it}}\boldsymbol{\tau} + \omega_i)$$

ω_i : Random scale effect to account for individual heterogeneity in variability of mood; allow covariation with ν_i

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

WS-variance for each subject at each occasion now has a log-normal distribution:

$$\sigma^2_{\epsilon_{it}} \sim \text{lognormal}(\boldsymbol{\omega}'_{\boldsymbol{it}}\boldsymbol{\tau}, \sigma^2_\omega)$$

Conditional variance of measurement y_{it} given covariates $\boldsymbol{\omega_{it}}$ of WS-variance sub-model:

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

Intraclass correlation:

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

Estimation of mixed effects location-scale model (1/2)

Standardizing random effects using Cholesky decomposition

We standardize the location and scale random effects to have a independent standard bivariate normal distribution:

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

$\boldsymbol{S}$ is the lower triangular Cholesky factor, where $\boldsymbol{S}\boldsymbol{S}^{\boldsymbol{T}} = \boldsymbol{\Sigma}_{\nu\omega}$. θ_{1i}, θ_{2i} are independent standard normals

We can re-write the mean and WS-variance sub-models using entries of the Cholesky factor

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + s_1\theta_{1i} + \epsilon_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \sigma_\nu\theta_{1i} + \epsilon_{it} \qquad \sigma_{\epsilon_{it}^2} = \exp(\boldsymbol{\omega}'_{it}\boldsymbol{\tau} + s_2\theta_{1i} + s_3\theta_{2i})$$

To further simplify the model, we can force $\sigma_{\nu\omega} = 0$ so that $s_2 = 0$ and $s_3 = \sigma_\omega$. Include θ_{1i} directly in WS-variance sub-model to model the relationship between random location effect and WS-variance

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

Parameters to be estimated: $\boldsymbol{\beta}, \boldsymbol{\tau}, \alpha_0, \boldsymbol{s} = [\tau_\ell, \tau_q, \sigma_\omega]$

Estimation of mixed effects location-scale model (2/2)

Maximum likelihood estimation

- Several software for implementing MLE of MELS mode
  - SAS PROC NLMIXED: Can be tweaked to accommodate estimation of MELS models. Tend to be slow
  - MIXREGLS: Standalone program, faster, accessible through R
  - MIXWILD: Standalone program, allows for the inclusion of multiple location and scale random effects

Bayesian approaches

- Produce very similar estimates to MLE when we use uninformative priors
- Choose diffuse priors on all parameters to be estimated (uniform and improper bounded uniform distribution). For random effects, choose standard bivariate normals by construction
- Can use WinBUGS, JAGS, and Stan (available through R and Python)
- Better at handling large number of random effects
  - Random slopes in the mean and WS-variance sub-models
  - 3-level models (when data are collected in distinct waves, can include wave-level random effects)

Simulation study: Inference for variance sub-model (1/2)

- The main feature of MELS is the inclusion of random scale effect in the joint-modeling of mean and variance. We conduct the following simulation study to compare the performance of models with and without random scale effect:

Data generation mechanism:

- 50 data sets, each with 150 subjects, measured at 25 occasions
- Same two covariates used in both mean and WS-variance sub-models were generated:
  - Binary, between-subject covariate: $X_i^{BS} \sim Bernouli(0.5)$
  - Continuous, within-subject covariate: $X_{it}^{WS} \sim N(0,1)$
- Set $\boldsymbol{\beta} = [1,1,1], \boldsymbol{\tau} = [0.5, 0.6, -0.4]$, $\sigma_\nu = 1$, $\sigma_\omega = 0.8$, $\sigma_{\nu\omega} = 0$
- Allow location effect to be log-linearly related to the WS-variance through parameter $\tau_\ell = 0.3$

Models fit:

- Models with and without random scale effect were fit, both models share the same mean sub-model

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

- Model 1 does not have scale effect. Since τ_ℓ models the relationship between random location and scale effect, it is also excluded in Model 1:

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

- Model 2 has random scale effect in WS-variance sub-model

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

- Both models fit using Bayesian approaches, implemented through Rstan using Hamiltonian Monte-Carlo aglorithm

Simulation study: Inference on variance sub-model (2/2)

Table 1: Model Estimates for 50 Simulated Data Sets

	Without Scale Effect			With Scale Effect		
Parameter	**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%

- Bias = Posterior mean – True value
- Average Interval Widths (AIW): 95% credible interval constructed from taking quantile values of posterior distribution
- Coverage: Used loosely here; however, Bayesian credible interval behaves similarly to frequentist confidence interval with sufficient sample and diffuse prior

- Both models give good estimates for $\boldsymbol{\beta}$ parameters
- For model without scale effect, estimates of WS-variance coefficients are biased and give insufficient coverage
- By not including scale effect, we are ignoring a major source of variation, which is individual differences in baseline variability of the response variable
- Coverage of the WS-variance sub-model parameters insufficient as we underestimate the uncertainty involved in their estimation, leading insufficient coverage and false positive findings

Simulation study: Robustness to serial correlation (1/2)

- So far, we have assumed that the within-subject error terms ϵ_{it} are independent across subjects and across occasions
- It is natural to suspect serial correlation of observations within the same subject
- Including lagged response variables $y_{i,t-1}, y_{i,t-2}$ ... does not work as it violates the assumption that random effects are independent from covariates
- The way to model this is to modify the covariance structure of ϵ_{it}, usually employing a AR(1) correlation structure
- Let ρ be the correlation coefficient between 0 and 1. For each subject i, specify a $T_i \times T_i$ correlation matrix P_i

$$P_i = \begin{bmatrix} 1 & \rho & \cdots & \rho^{T_i-1} \\ \rho & 1 & \cdots & \rho^{T_i-2} \\ \vdots & & & \vdots \\ \rho^{T_i-1} & \rho^{T_i-2} & \cdots & 1 \end{bmatrix} \qquad \Sigma_{\epsilon_i} = \begin{bmatrix} \sigma^2_{\epsilon_{i,1}} & \sigma_{\epsilon_{i,1}}\sigma_{\epsilon_{i,2}} & \cdots & \sigma_{\epsilon_{i,1}}\sigma_{\epsilon_{i,T_i}} \\ \sigma_{\epsilon_{i,2}}\sigma_{\epsilon_{i,1}} & \sigma^2_{\epsilon_{i,2}} & \cdots & \sigma_{\epsilon_{i,2}}\sigma_{\epsilon_{i,T_i}} \\ \vdots & & & \vdots \\ \sigma_{\epsilon_{i,T_i}}\sigma_{\epsilon_{i,1}} & \sigma_{\epsilon_{i,T_i}}\sigma_{\epsilon_{i,2}} & \cdots & \sigma^2_{\epsilon_{i,T_i}} \end{bmatrix} \odot \begin{bmatrix} 1 & \rho & \cdots & \rho^{T_i-1} \\ \rho & 1 & \cdots & \rho^{T_i-2} \\ \vdots & & & \vdots \\ \rho^{T_i-1} & \rho^{T_i-2} & \cdots & 1 \end{bmatrix}$$

- The within-subject variance-covariance matrix Σ_{ϵ_i} accounting for serial correlation can then be specified as above
- Currently, standard software for estimating MELS such as MIXREGLS and MIXWILD do not have the option to specify correlation structure
- For our simulation, we modify the data generation mechanism of the previous simulation to include auto-correlation, with $\rho = 0.7$, and fit a MELS model ignoring the autocorrelation structure to test robustness of model estimates

Simulation study: Robustness to serial correlation (2/2)

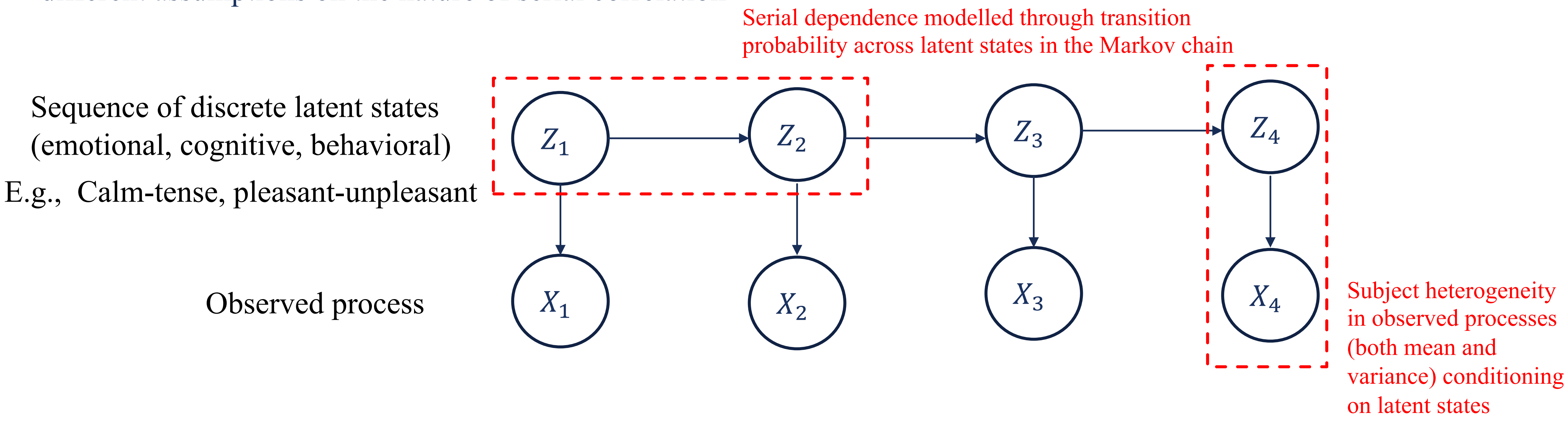
Table 2: Model Estimates for 50 Simulated Data Sets

	Ignoring AR(1)		
Parameter	**Bias**	**AIW**	**Coverage**
$\beta_{intercept} = 1$	-0.002	0.613	96%
$\beta_{BS} = 1$	-0.042	0.864	98%
$\beta_{WS} = 1$	-0.004	0.125	94%
$\tau_{intercept} = 0.5$	1.142	0.456	0%
$\tau_{BS} = 0.6$	-0.568	0.645	0%
$\tau_{WS} =$ -0.4	0.407	0.103	0%
$\sigma_{\nu} = 1$	0.235	0.348	0.16%
$\sigma_{\omega} = 0.8$	0.147	0.241	0.32%
$\tau_{\ell} = 0.3$	-0.287	0.37	0.18%

- Estimates for the mean sub-model is robust, but estimates and inference for the WS-variance sub-model no longer valid
- Strong auto-correlation make WS errors look more homogenous across covariate values, so estimates of coefficients on regressors are biased towards 0, and intercept is biased upward
- Estimates of BS variance biased upwards, results in mis-leading intraclass-correlation-coefficient (ICC)
- In practice, serial correlation can be more complicated
- Fixes: modeling changes in response variable across observations, including lagged terms of regressors

Mixed location-scale hidden Markov model

- Alternative framework for dealing with serial correlation, does not model data generated under AR(1) structure, puts different assumptions on the nature of serial correlation



- Assumptions not necessarily applicable to all psychological and clinical processes, but when they do, is a useful framework to model the different levels of variability in EMA data while accounting for serial dependence of observations

Formulation of mixed location-scale HMM

- Let there be $i = 1,2, \ldots, N$ subjects, each with T_i observations taken over equally spaced time intervals
- Assume there are two latent states, let $Z_i = (z_1, z_2, \ldots, z_{T_i})$ be the sequence of latent states subject i goes through, which follow a first order Markov chain with transition probability matrix A and initial probability distribution: $\pi_1, \pi_2 = 1 - \pi_1$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \; a_{11} + a_{12} = 1, \; a_{21} + a_{22} = 1$$

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

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

- Reflect individual heterogeneity in the observed process as below:

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

  $\boldsymbol{x_{it}}, \boldsymbol{w_{it}}$: Regressors for mean and WS-variance sub-models

  $\boldsymbol{\beta_k}, \boldsymbol{\alpha_k}$: Corresponding coefficients, different across latent states

  ν_i, ω_i : Location and scale random effects, follows a bivariate normal distribution as specified in MELS

  Parameters to be estimated: $(A, \pi_1, \boldsymbol{\beta_k}, \boldsymbol{\alpha_k}$ for $k = 1,2)$

- Model can be estimated using Bayesian approach combined with backward-forward algorithm for HMM

Simulation study on mixed location-scale HMM (1/2)

- We conduct a simulation study to verify the validity of our estimation procedure as well as to examine the properties of mixed location-scale HMM

Data generating mechanism:

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 mixed location-scale HMM

Simulation study on mixed location-scale HMM (2/2)

Table 3: Model estimated for 50 data sets

	HMM			Mixed HMM			Mixed location-scale HMM		
Parameter	**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%

- Mixed location-scale HMM is the only model that provides sufficient coverage and valid estimates; successfully identifies the low-mean, high-variance state and the high-mean, low-variance state
- For mixed HMM ignoring scale effect, estimates for $\sigma_{\epsilon,1}$ and $\sigma_{\epsilon,2}$ more separate than they actually are. HMM attribute more of the total variation to differences across states when they should be attributed to individual heterogeneity
- Likewise for μ_1 and μ_2 when we ignore location random effect

Data analysis example (1/4)

- **Data:** Random prompts portion of motivating study
  - 17,402 random prompts obtained from 510 students, averaging 34 prompts per student (range = 3 to 58)
- **Variables:**
  - *Negative affect:* Response variable
  - *Others:* Subject with others at time of measurement
    - $Others_{BS}$: For subject i, proportion of time subject i is with others, $\overline{Others_i}$
    - $Others_{WS}$: $Others_{it} - \overline{Others_i}$
  - *Genderf:* Subject is female or not
  - *Age15:* Subject age minus 15 (average age)
  - *COPEc:* Grand-mean centered version of a coping scale measurement at baseline
  - *Time indicators:* 4 indicators to indicate time block of measurement
- Precise timestamps not available

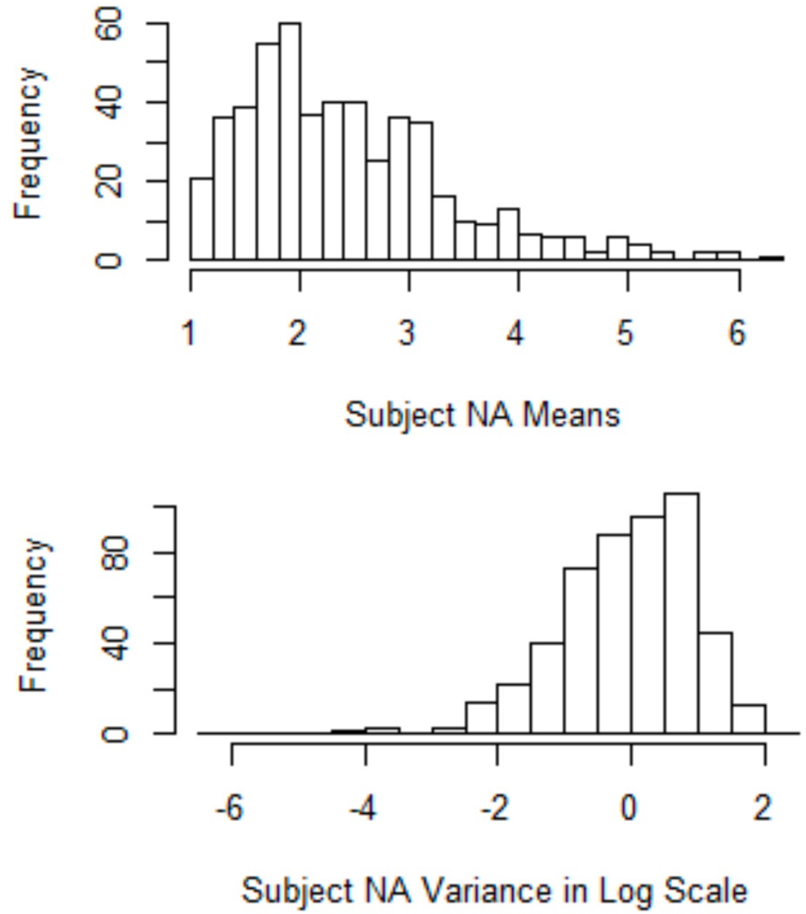


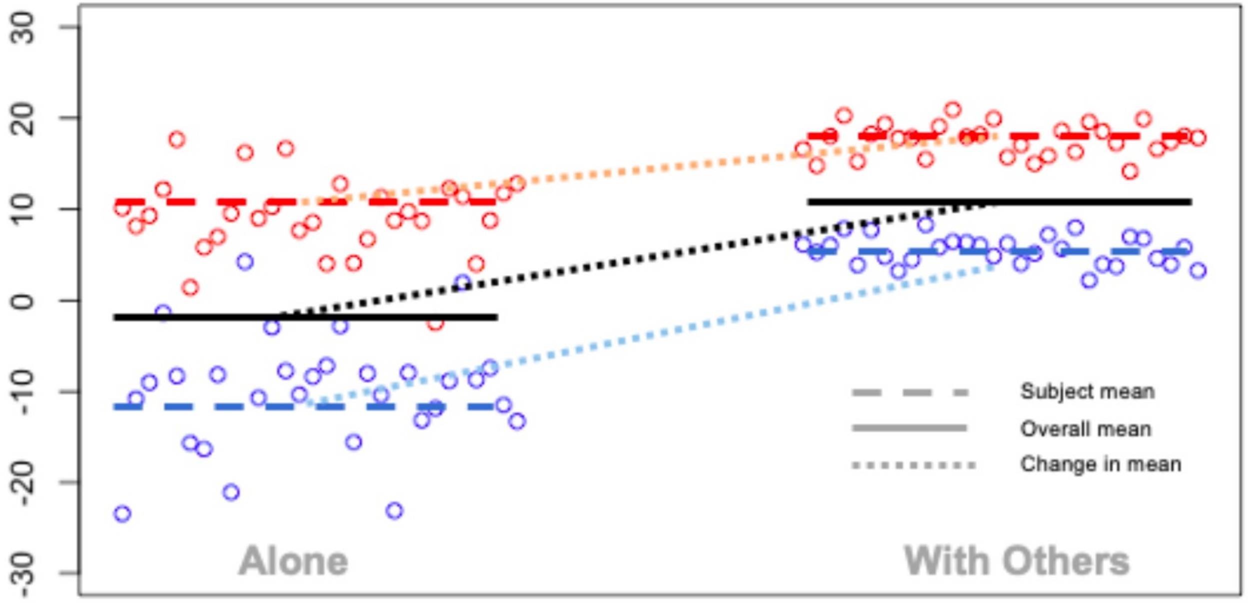
Figure 1: Histograms of subject means and variance for *negative affect*

Data analysis example (2/4)

- For this dataset, we can consider modeling BS-variance in terms of covariates as well. Let $\boldsymbol{u}_{it}$ be the regressors and $\boldsymbol{\alpha}$ be the corresponding coefficients, the log-linear model is:

$$\sigma^2_{\nu_{it}} = \exp(\boldsymbol{u}'_{it}\boldsymbol{\alpha})$$

- Effect of subject-level covariates such as gender: Are girls more similar to each other than boys are?
- Effect of occasion-level covariates:
  - Occasion-level covariates can have effects on both WS and BS variance



- Hypothesized relationship (not from real data)
  - Red and blue represent two subjects
  - Dotted bar: Subject means at each state
  - Solid bar: Overall mean at each state
  - WS variance: Dispersion of dots around subject mean
  - BS variance: Gap between subject means

Data analysis example (3/4)

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

Variable	**Estimate**	**Std Error**	**z-value**	**p-value**
$\boldsymbol{\beta}$ *:Regression Coefficients*				
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
$\boldsymbol{\alpha}$ *:BS Variance parameters: log-linear model*				
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

- Both WS and BS version of *Others* have significant negative effect. Higher overall tendency to be with others as well as momentary effect of being with others can both dampen negative affect
- Higher coping score associated with lower negative affect
- When subjects are with others, there are less heterogeneity in mean level of negative affect across subjects
- Subjects with higher coping scores are also less heterogenous

Data analysis example (4/4)

τ :WS variance parameters: log-linear model				
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
Random Location (Mean) Effect on WS Variance				
Linear Effect: τ_ℓ	0.894	0.052	17.234	0.000
Quad Effect: τ_q	-0.271	0.031	-8.856	0.000
Random Scale Standard Deviation				
σ_ω	0.621	0.026	24.172	0.000

- Being with others associated with more stable negative affect level. Having a higher coping score also associated with more stable negative affect level
- Higher location random effect also contributes to higher WS-variance
- For full data set, can analyze the effect of smoking
- No time-stamps available, so could not fit models accounting for temporal dependencies

Modeling time-varying effects in EMA data

A motivating study

EMA study on smoking-cessation therapies from Lanza et al. (2014)

- **Participants**: 1,504 smokers (58% female) who smoked at least 10 cigarettes a day for 6 months and were motivated to quit smoking
  - Randomized to monotherapy, combination therapy, and placebo
- **Data collection:** Each participant had a target quit day (TQD) after receiving treatment or placebo
  - For the 2 weeks prior and 2 weeks post target quit date, four daily EMA measurements (after waking, before sleep, and the other 2 at random times) collected
  - EMA prompts assessed smoking urge (series of items rating severity of withdrawal symptoms), negative affect of subject within last 15 minutes
  - Covariate: Baseline nicotine dependence measured before study
- Only used data from 1,106 participants who succeeded in establishing initial abstinence (no smoking for at least 24hr after target quit day) and did not relapse (defined as smoking on 7 consecutive days) during the first 2 weeks following TQD (29,497 EMA entries)

Questions of Interest:

1. What is the association between smoking urge (response variable) and negative affect / baseline nicotine dependence over a quit attempt? How does it change over time after TQD, and how do they differ across treatment groups?

2. Are there any points in time at which effects of treatment changes?

Formulation of time-varying effects model

Let y_{it} denote the measurement of response variable for subject i $(i = 1,2, \dots, N)$ on occasion t $(t = 1,2, \dots, T_i)$

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

t_{it}: Time stamp of subject i's t-th observation, recorded in relation to some reference time point

x_{it}: Some regressor of interest, could be time varying or non time-varying

ϵ_{it}: Error terms. Can assume it to be independent and normally distributed, can also place some serial correlation structure on it

$\beta_0(\cdot), \beta_1(\cdot)$: Intercept and coefficient for the regressor of interest; continuous function of time, no parametric assumptions

Can be generalized easily. Let $g(\cdot)$ be the link function. The generalized TVEM is then:

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

Estimation of time-varying effects model (1/2)

Estimate $\beta(\cdot)$ using P-spline (Penalized-splines) with truncated-power basis:

- Assume data were collected over time interval [a, b], divide [a,b] into K+1 intervals, K is a tuning parameter
- Place knots $\{\tau_1, \tau_2, \dots, \tau_k\}$ (τ_0=a, τ_{k+1}=b) to divide into K+1 intervals, placement of knots is a tuning parameter
- Within each of these intervals, estimate coefficient functions locally with low-order polynomials formed by linear combinations of truncated power basis of order q, q is a tuning parameter (usually set to 2 or 3)
- Truncated power basis of order q has the form:

$$1, t, t^2, \dots, t^q, (t-\tau_1)_+^q, (t-\tau_2)_+^q, \dots, (t-\tau_k)_+^q \qquad (t-\tau)_+^q = \begin{cases} 0, & \text{if } t \le \tau \\ (t-\tau)^q, & \text{otherwise} \end{cases}$$

Say we pick q=2, the coefficient functions can then be approximated as:

$$\beta_0(t) \approx a_0 + a_1 t + a_2 t^2 + \sum_{k=1}^{K} a_{2+k}(t-\tau_k)_+^2 \qquad \beta_1(t) \approx b_0 + b_1 t + b_2 t^2 + \sum_{k=1}^{K} b_{2+k}(t-\tau_k)_+^2$$

The regression model is now a model with 2K+6 covariates in total:

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

Estimation of time-varying effects model (2/2)

To avoid overfitting, we smooth the estimated coefficient functions using penalty terms. There are two approaches.

First approach is to penalize coefficients for truncated power basis: $\{a_{2+j}, j = 1,2 \ldots, K\}$ and $\{b_{2+j}, j = 1,2 \ldots, K\}$, since these coefficients represent “jumps” that contributes to model complexity

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

λ_1, λ_2 are tuning parameters selected using generalized cross-validation and similar algorithms

Alternative approach (Krivobokova & Kauermann 2007) treats the coefficients for truncated power basis as random effects with normal distribution, each with variance η_1 and η_2. If η big, coefficients deviate largely from 0, so lots of jumps

$$a_{2+k} \sim N(0, \eta_1), k = 1, 2, \ldots, K$$
$$b_{2+k} \sim N(0, \eta_2), k = 1, 2, \ldots, K$$

Random coefficient effects for truncated power functions estimated using empirical bayes method. Krivobokova & Kauermann have shown that when we use REML to estimate the parameters, the result is comparable to the first approach

Simulation study: Capturing non-linear curves (1/4)

We conduct simulation study to examine TVEM's capability in capturing coefficient functions that are non-linear and cyclical over time

Data generation mechanism:

- Generated two data sets, with N=50 and N=200 respectively
- Number of observations per individual is set to 30 for both data sets
- Measurement time for each observations distributed uniformly over [0,1]
- Two time-varying covariates $x_{1it} \sim Bernouli(0.5)$ and $x_{2it} \sim N(0,1)$ are included
- The model is specified as: $y_{it} = \beta_0(t_{it}) + \beta_1(t_{it})x_{1it} + \beta_2(t_{it})x_{2it} + e_{it}$
- With coefficient functions specified as:

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

- The error terms follow an AR(1) structure: $e_{it} = \rho e_{i(j-1)} + \epsilon_{it},\ \epsilon_{it} \sim N(0,1)$
- We use TVEM SAS macro provided by Li et al (2017), which uses P-spline
- For simplicity, set the order of truncated power basis to be 3, and number of knots to be 10

Simulation study: Capturing non-linear curves (2/4)

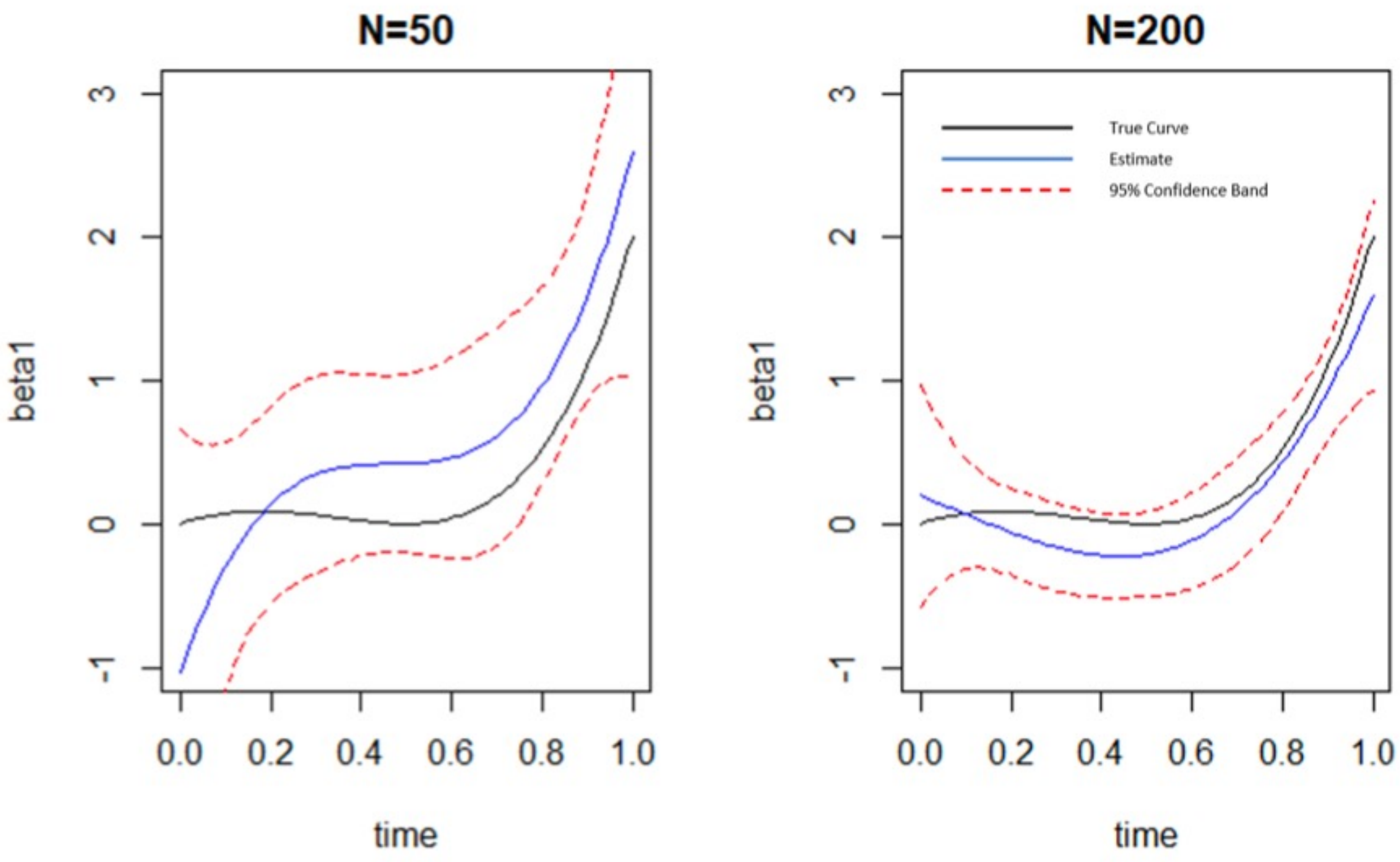


Figure 5: Estimate for $\beta_0(t)$ from first simulation study

Simulation study: Capturing non-linear curves (3/4)

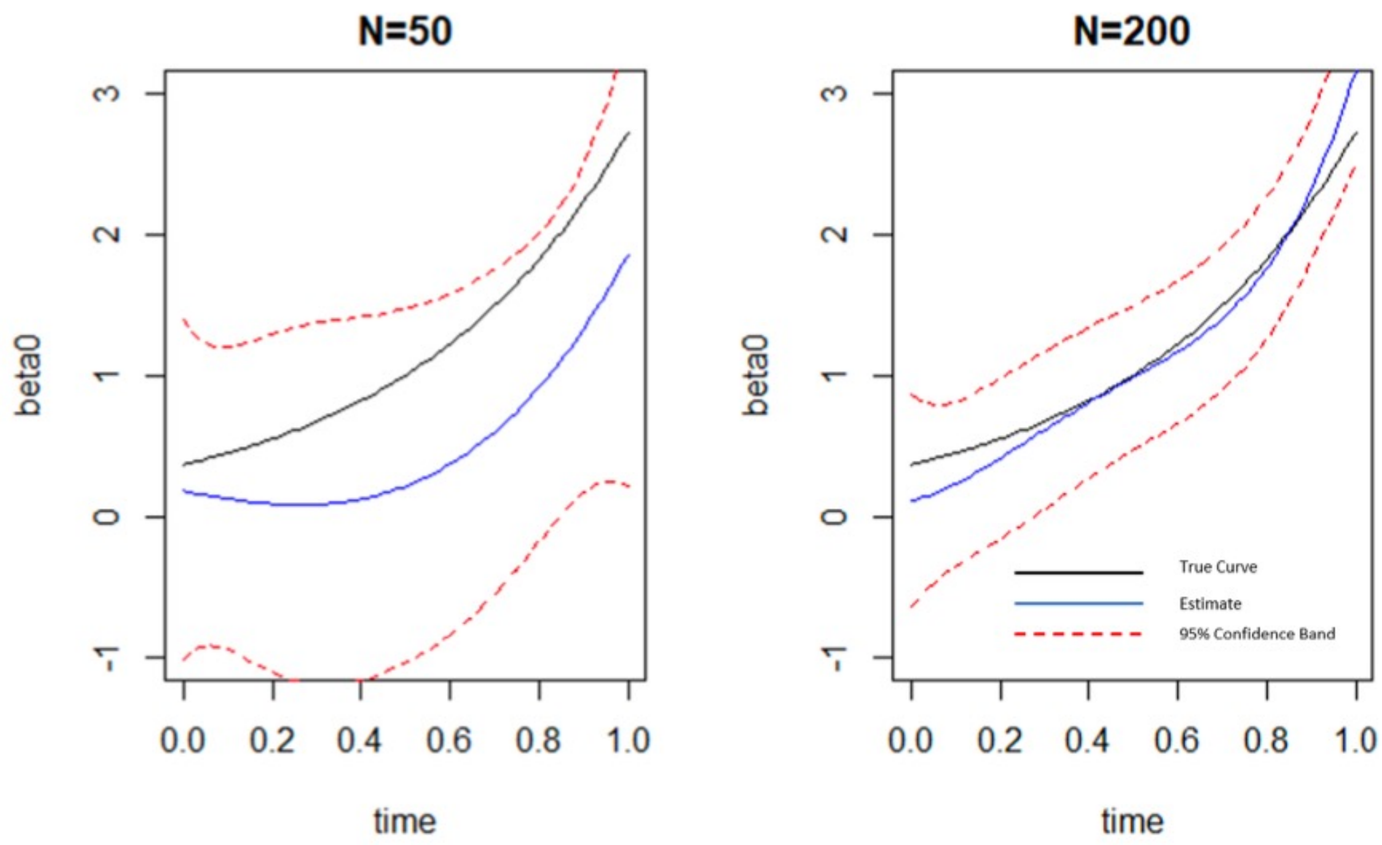


Figure 6: Estimate for $\beta_1(t)$ from first simulation study

Simulation study: Capturing non-linear curves (4/4)

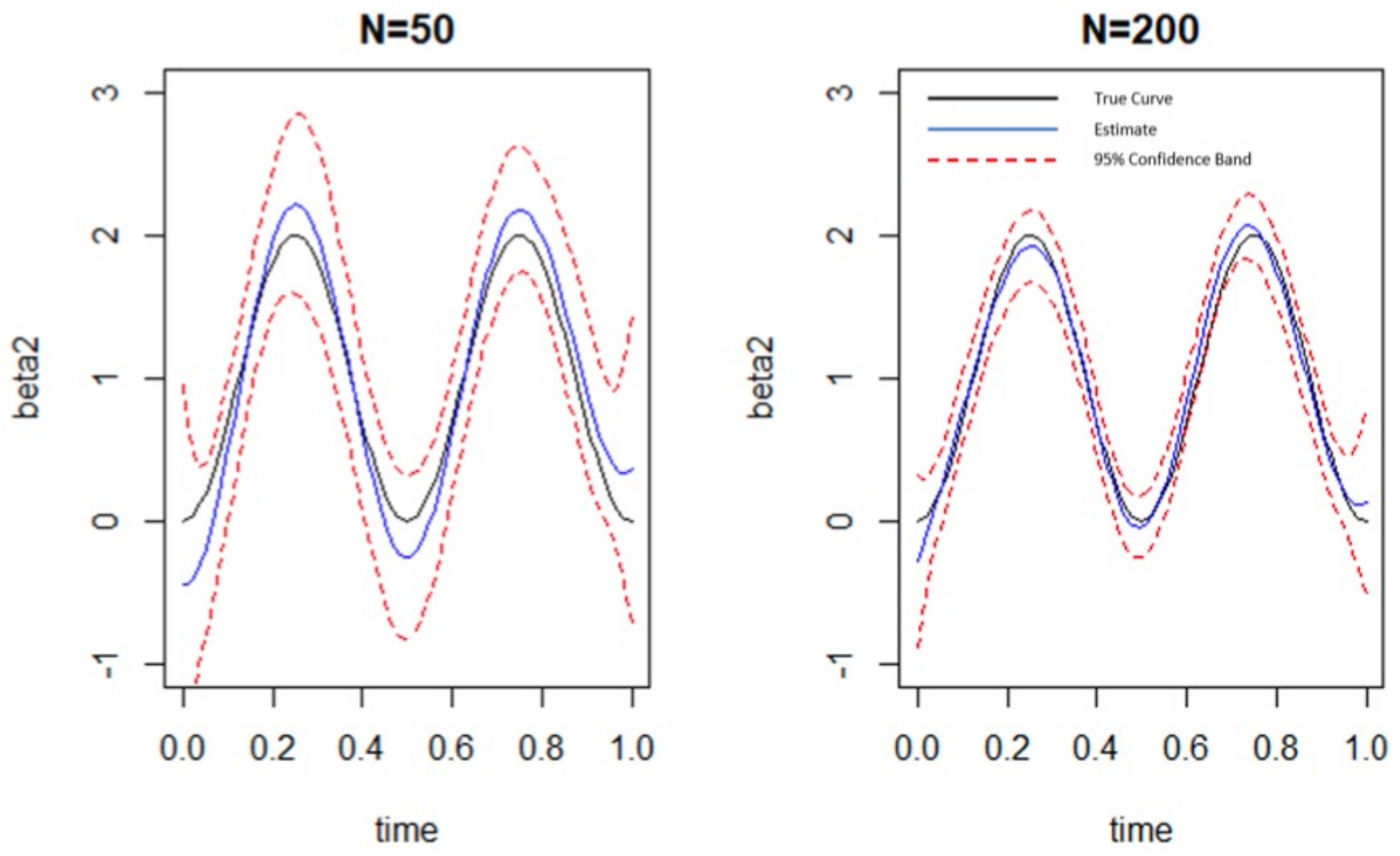


Figure 7: Estimate for $\beta_2(t)$ from first simulation study

Discussion on TVEM

Data analysis example:

- For the motivating study on smoking-cessation therapies, 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 to compare the shape of coefficient functions over the first 2 weeks post-quit
  - Study found that association between negative affect and craving in treatment groups initially weakened significantly relative to placebo group, but vanished after day 2, whereby negative affect and craving had a positive association
    - Identifies potential intervention windows to make treatments more effective
  - Association between baseline nicotine dependence and craving among combination therapy group was significantly lowered compared to other groups after Day 6
  - Only models dynamics associations, unable to tease apart directions of relationship (e.g., does negative affect increase smoking urge, or inability to satisfy smoking urge increases negative affect?)

Extensions:

- Unlike MELS, does not focus on individual heterogeneity as much. Modeling coefficient curve for each individual unnecessary as we are more interested in shapes
- Mixture Time Varying Effects Model (Li et al 2017) assigns subjects to latent groups, members of the same group share similarly shaped coefficient functions

Discussions

Discussions

- Both MELS and TVEM are developed to unlock the contextual and temporal information embedded in EMA data
- A common draw-back is that some variations of these models are computationally burdensome
- In the future, EMA data could be increasingly integrated with mHealth (mobile health) analytic platforms; effectively integration with other data streams could extend the range of questions EMA data can answer
- Ultimate 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
- Traditional models such as MELS and TVEM could assist in building these predictive intervention models
  - Subject's dominant latent state identified by mixed location-scale hidden Markov model could inform development of individually tailored interventions
  - Time-varying effect models could identify key time frames at intervention is most needed through the shape of its coefficient curves