Shared Parameter Mixed-Effects Location Scale Models for Intensive Longitudinal Data

Donald Hedeker
Department of Public Health Sciences
University of Chicago

hedeker@uchicago.edu
https://voices.uchicago.edu/hedeker/

New York Metro Area Chapter of the American Statistical Association
September 16, 2022

Supported by NCI grant R01 CA240713 (Hedeker & Dunton) and by NCI and FDA Center for Tobacco Products (CTP) grant R01 CA184681 (Mermelstein, PI).

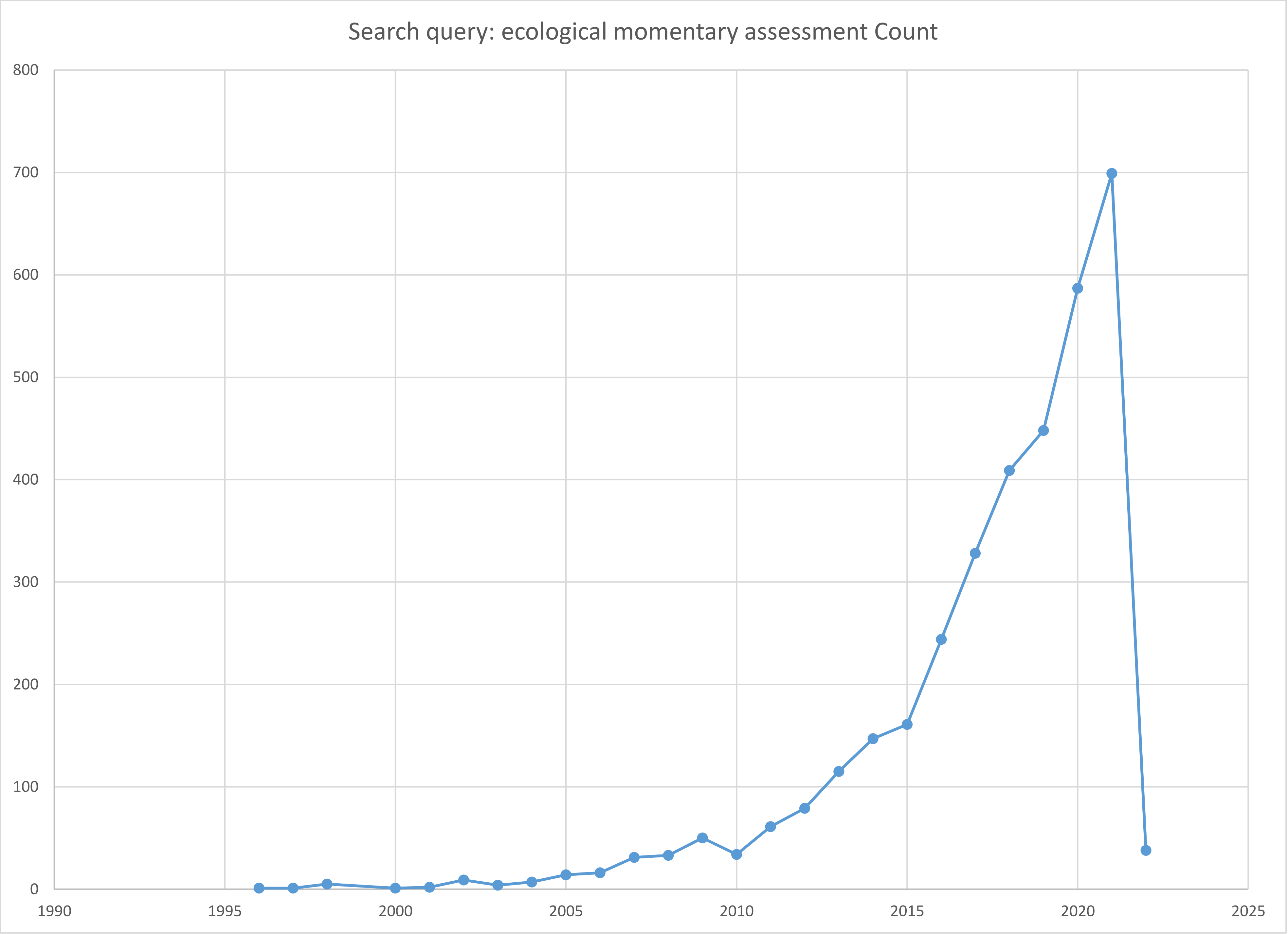
Ecological Momentary Assessment (EMA) data

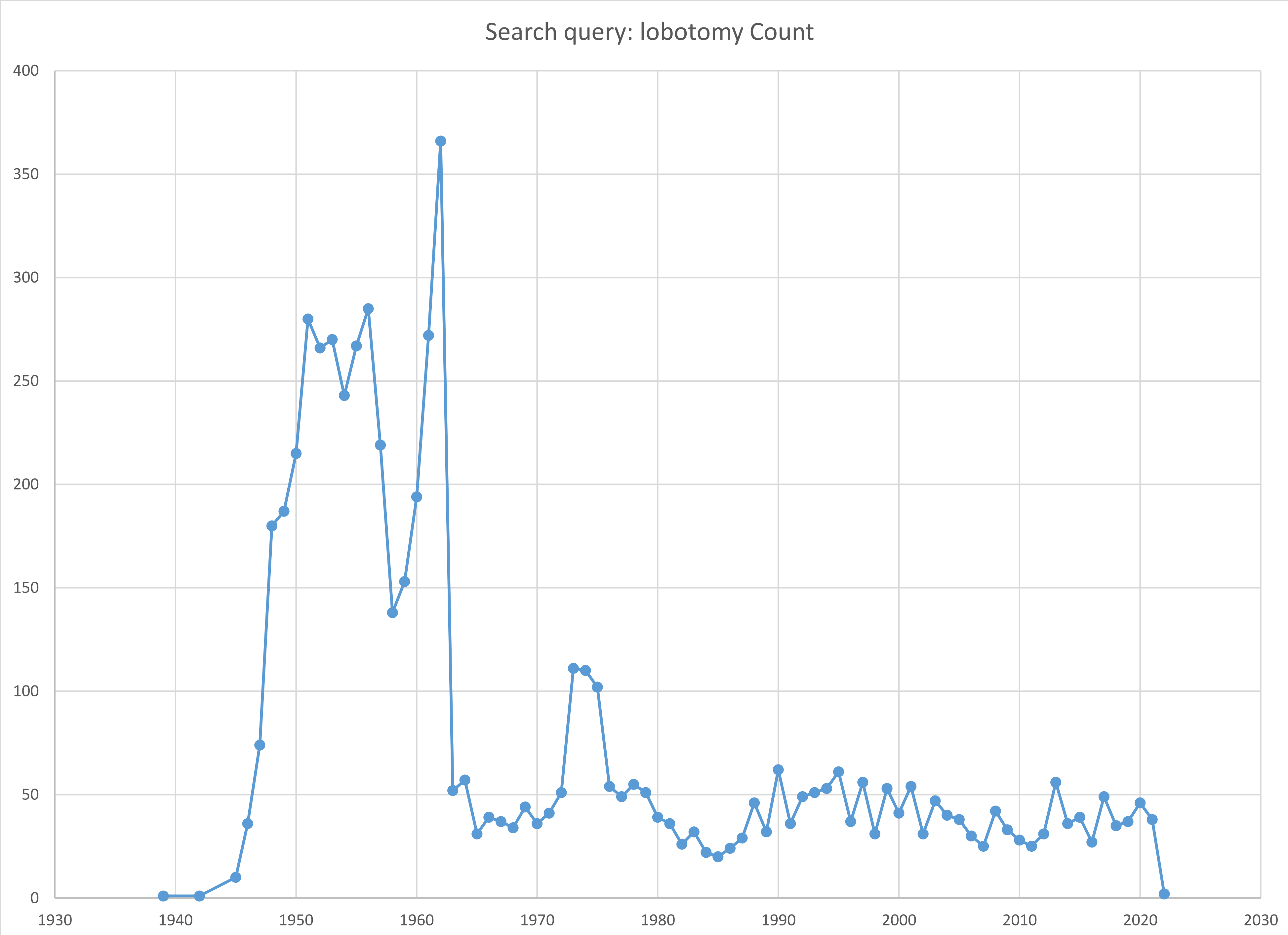
experience sampling and diary methods, intensive longitudinal data

- Subjects provide frequent reports on events and experiences of their daily lives (*e.g.*, 30-40 responses per subject collected over the course of a week or so)
  - electronic diaries: palm pilots, personal digital assistants (PDAs), smart phones

- Capture particulars of experience in a way not possible with more traditional designs
  e.g., allow investigation of phenomena as they happen over time

- Reports could be time-based, following a fixed-schedule, randomly triggered, event-triggered


Search query: ecological momentary assessment Count
800
700
600
500
400
300
200
100
0
1990
1995
2000
2005
2010
2015
2020
2025


Search query: lobotomy Count
400
350
300
250
200
150
100
50
0
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030

Data are rich and offer many modeling possibilities!

- person-level and occasion-level determinants of occasion-level responses $\Rightarrow$ potential influence of context and/or environment
  e.g., subject response might vary when alone vs with others

- allows examination of why subjects differ in variability in addition to mean level
  - within-subjects (WS) variance
    e.g., subject degree of stability/inconsistency could vary by gender or age

Carroll (2003) Variances are not always nuisance parameters, *Biometrics*.

Submodel One: Mixed-effects location-scale model for measurement y of subject i ($i = 1, 2, \ldots, N$) on occasion j ($j = 1, 2, \ldots, n_i$)

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i + \epsilon_{ij}$$

$\boldsymbol{x}_{ij} = p \times 1$ vector of regressors (including a column of ones)

$\boldsymbol{\beta} = p \times 1$ vector of regression coefficients

$\boldsymbol{z}_{ij} = r \times 1$ vector of random effect variables (including a column of ones)

$\boldsymbol{v}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\upsilon})$ BS variance; how homogeneous/heterogeneous are subjects?

$\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon})$ WS variance; how consistent/erratic are the data within subjects?

Log-linear model for WS variance (exp function ensures a positive multiplicative factor, and so resulting variance is positive)

$$\sigma^2_{\epsilon_{ij}} \;=\; \exp(\boldsymbol{w}'_{ij}\boldsymbol{\tau} + \boldsymbol{\upsilon}_i\boldsymbol{\tau}^* + \omega_i) \qquad \text{where} \qquad \omega_i \sim N(0, \sigma^2_\omega)$$

$$\log(\sigma^2_{\epsilon_{ij}}) \;=\; \boldsymbol{w}'_{ij}\boldsymbol{\tau} + \boldsymbol{\upsilon}_i\boldsymbol{\tau}^* + \omega_i$$

- ω_i are log-normal subject-specific perturbations of WS variance; "scale" random effects - how does a subject differ in terms of the variation in their data
- $\boldsymbol{\upsilon}_i$ are "location" random effects - how does a subject differ in terms of the mean of their data (e.g., intercepts and slopes)
- $\boldsymbol{\tau}^*$ are association parameters between location and scale

Parameters of Submodel One:

- Fixed effects on the mean: $\boldsymbol{\beta}$
- Random subject effects on the mean (location): $\boldsymbol{\upsilon}_i$ with variance-covariance matrix $\boldsymbol{\Sigma}_{\upsilon}$
- Fixed effects on the WS variance: $\boldsymbol{\tau}$
- Random subject effect on the WS variance (scale): ω_i with variance σ^2_{ω}
- Association of subject's location and scale $\boldsymbol{\tau}^*$

Nordgren, R., Hedeker, D., Dunton, G., & Yang, C.-H. (2020). Extending the mixed-effects model to consider within-subject variance for Ecological Momentary Assessment data. *Statistics in Medicine, 39*:577-590.

Submodel Two: Submodel One random effects (e.g., intercept υ_{0i}, slope υ_{1i}, scale ω_i) are shared, in addition to subject-level variables $\boldsymbol{x}_i^{(2)}$, as predictors of subject-level outcome $y_i^{(2)}$

$$y_i^{(2)} = \alpha_0 + \alpha_1 \upsilon_{0i} + \alpha_2 \upsilon_{1i} + \alpha_3 \omega_i + \boldsymbol{x}_i^{(2)} \boldsymbol{\alpha} + \varepsilon_i$$

- Can include interactions of random effects with each other and/or other subject-level variables $\boldsymbol{x}_i^{(2)}$ (e.g., location effect varies depending on erraticism/consistency of a subject)

Observational Study of Dual Users (Mermelstein)

- Ongoing observational study of "dual users" who are primarily combustible cigarette smokers who are "early" in their trial/uptake of e-cigarettes
- Goals: examine patterns and predictors of changes in tobacco use patterns (combustible/e-cigarette patterns), with an emphasis on subjective experience and contexts of use of products, over a 12 month time frame
  - EMA at multiple time points (random prompts, Cig & ECig events)
  - Biweekly reports of cigarette and e-cigarette use over the 12 months (after baseline EMA)
  - Extensive psychosocial and behavioral questionnaires at baseline and 12 months
  - Present data on 240 dual users who completed 12 months and provided at least 2 Cig and 2 ECig event reports during EMA

Outcomes - $N = 240$ subjects

- Submodel 1: satisfaction attributed to Cig & ECig events (EMA)
  - How satisfying was tobacco product used?
  - How pleasurable was tobacco product used?
  - each rated on 1 (not al all) to 10 (very much) scale
  - average = reported satisfaction attributed to cig & ecig use
  - 5339 cig events and 2510 ecig events (total = 7849 events)

- submodel 2: subject average of daily cigarette smoking rate (daily ecig rate) from biweekly reports (post-EMA)

Submodel 1: satisfaction attributed to Cig & ECig events (EMA)

$$\texttt{Satis}_{ij} = \beta_0 + \beta_1 \texttt{Ecig}_{ij} + \upsilon_{0i} + \upsilon_{1i} \texttt{Ecig}_{ij} + \varepsilon_{ij}$$

Cholesky of random effects var-covar matrix, $\mathbf{\Sigma}_\upsilon = \boldsymbol{S}\boldsymbol{S}'$; $\boldsymbol{\upsilon}_i = \boldsymbol{S}\boldsymbol{\theta}_i$, where $\boldsymbol{\theta}_i$ = std normal random effects

$$\texttt{Satis}_{ij} = \beta_0 + \beta_1 \texttt{Ecig}_{ij} + (s_0 + s_{01} \texttt{Ecig}_{ij})\, \theta_{0i} + s_1 \texttt{Ecig}_{ij}\, \theta_{1i} + \varepsilon_{ij}$$

$$\sigma^2_{\epsilon_{ij}} \;=\; \exp(\tau_0 + \tau_1 \texttt{Ecig}_{ij} + \tau_0^* \theta_{0i} + \tau_1^* \theta_{1i} + \sigma_\omega \theta_{2i})$$

Submodel 2: Average daily smoking rate from biweekly reports (over the 12 months)

$$\sqrt{\texttt{SmkRate}}_i = \alpha_0 + \alpha_1 \theta_{0i} + \alpha_2 \theta_{1i} + \alpha_3 \theta_{2i} + \varepsilon_i$$

SAS PROC NLMIXED: likelihood estimation for shared parameter location-scale mixed model (models share random efffects `theta0`, `theta1`, `theta2`)

```
PROC NLMIXED DATA=all2 noad qpoints=11;
PARMS b0=7.38 bEcig=.10 t0=.335 tEcig=-.12
      s0=1.6 s01=0 s1=1.1  ttheta0=0 ttheta1=0 sdscale=.05
      a0=2.37 aCig=0 aEcig=0  aScale=0 vares=1;
lla=0;llb=0;   pi = arcos(-1);
if (ind eq 0) then
do;
  mu = b0 + bEcig*Ecig + (s0 + s01*Ecig)*theta0 + s1*Ecig*theta1;
  vare = exp(t0 + tEcig*Ecig + ttheta0*theta0 + ttheta1*theta1 + sdscale*theta2);
  lla = log(1 / (SQRT(2*pi*vare))) + (-(outcome-mu)**2)/ (2*vare);
end;
if (ind eq 1) then
do;
  mus = a0 + aCig*theta0 + aEcig*theta1 + aScale*theta2;
  llb = log(1 / (SQRT(2*pi*vares))) + (-(outcome-mus)**2)/ (2*vares);
end;
ll = lla+llb;
MODEL outcome ~ GENERAL(ll);
RANDOM theta0 theta1 theta2 ~ NORMAL([0,0,0], [1,0,1,0,0,1]) SUBJECT=id;
RUN;
```

Joint Model: Satisfaction & Biweekly Smoking Rate mean: 240 subjects with 7849 events (5339 Cig & 2510 Ecig)

Parameter	Estimate	Std Error	t Value	Pr > |t|
BETA (regression coefficients)				
b0	7.4601	0.02170	343.80	<.0001
bEcig	-0.2216	0.04554	-4.87	<.0001
TAU (WS variance parameters: log-linear model)				
t0	0.4632	0.05053	9.17	<.0001
tEcig	-0.1701	0.04386	-3.88	0.0001
Cholesky of Random (location) Effect Variances and Covariances				
s0	1.3322	0.01200	111.02	<.0001
s01	-0.5310	0.02144	-24.76	<.0001
s1	1.2961	0.02007	64.57	<.0001
Random location effects on WS variance (log-linear model)				
ttheta0	-0.2263	0.02825	-8.01	<.0001
ttheta1	0.07560	0.03052	2.48	0.0139
Random scale standard deviation				
sdscale	0.9108	0.02974	30.62	<.0001

```
Regression model for  sr_CigRate_mean

a0              2.4120    0.07451     32.37     <.0001
aCig            0.1593    0.06314      2.52     0.0123
aEcig          -0.3596    0.07662     -4.69     <.0001
aScale         0.09194    0.07223      1.27     0.2043

error var       1.2222     0.1145     10.67     <.0001
```

- Both submodels share random subject intercept (**aCig**), slope (**aEcig**), and scale (**aScale**)
- Higher levels of random intercept (satisfaction from Cigs) and lower slopes (diff between Ecigs and Cigs) lead to increased smoking rates

- Shared parameter model concurrently models data from EMA (satisfaction from smoking/ecig events) and post-EMA (average biweekly cigarette smoking/ecig rate)
- Both models share random effects (intercept, slope, scale)
- Post-EMA data can influence estimation of random effects (variance-covariance matrix of random effects)
- Timewarp problem?
- Maybe better to fit two models separately in two stages

Stage 1 analysis - EMA Satisfaction ratings (N=240; 7849 events)

Variable	Estimate	Std Error	z-value	p-value
BETA (regression coefficients)				
Intercept	7.37402	0.09956	74.06853	0.00000
Ecig	0.09794	0.10264	0.95421	0.33998
Random (location) Effect Variances and Covariances (corr = 0.496)				
Intercept	2.51752	0.23670	10.63607	0.00000
Covariance12	-1.18002	0.18754	-6.29218	0.00000
Ecig	2.24915	0.21864	10.28713	0.00000
TAU (WS variance parameters: log-linear model)				
Intercept	0.33527	0.06862	4.88602	0.00000
Ecig	-0.12278	0.04475	-2.74375	0.00607
Random location effects on WS variance (log-linear model)				
Intercept	-0.37459	0.06678	-5.60945	0.00000
Ecig	0.02434	0.06876	0.35394	0.72338
Random scale standard deviation				
Std Dev	0.92704	0.04689	19.76850	0.00000

Stage 2 analysis

Stage 1 random subject effect estimates (e.g., intercept $\hat{\theta}_{0i}$, slope $\hat{\theta}_{1i}$, scale $\hat{\theta}_{2i}$) and other subject-level variables $\boldsymbol{x}_i^{(2)}$ used as regressors to predict a Stage 2 subject-level outcome $y_i^{(2)}$

$$y_i^{(2)} = \alpha_0 + \alpha_1 \hat{\theta}_{0i} + \alpha_2 \hat{\theta}_{1i} + \alpha_3 \hat{\theta}_{2i} + \boldsymbol{x}_i^{(2)} \boldsymbol{\alpha} + \varepsilon_i$$

- Random subject effects are EB estimates with estimated uncertainty; “plausible value” replications of the the random effects can be performed (Mislevy, 1991) with repeated samples from each subject’s posterior distribution
- EB estimates are biased toward zero, though this is mitigated if large n_i, large random effects variance, small residual variance (Liu, Kuppens, & Bringmann, 2019)

Stage 2 analysis - N = 240 subjects - Average Daily Cig Rate

```
Descriptives
------------

 Dependent variable
                           mean         min         max     std dev
 ------------------------------------------------------------------
 sr_CigRate_mean         2.3363      0.1048      5.9161      1.1900


 Random Location and Scale EB mean estimates
                           mean         min         max     std dev
 ------------------------------------------------------------------
 Locat_1                 0.0000     -2.5639      1.6546      0.9706
 Locat_2                -0.0000     -5.1979      3.4499      0.9170
 Scale                  -0.0000     -4.1966      1.7523      0.9481
```

Stage 2 analysis - averaged regression results from 500 plausible value replications of Stage 1 random effects

Variable	Estimate	Std Error	z-value	p-value
Intercept	2.33637	0.07343	31.81973	0.00000
Stage 1 Intercept	0.18206	0.07592	2.39790	0.01649
Stage 1 Slope	-0.31452	0.07893	-3.98476	0.00007
Stage 1 Scale	0.10463	0.07660	1.36601	0.17194
Residual Variance	1.26680	0.11773	10.76049	0.00000

$\Rightarrow$ higher levels of Stage 1 intercept (satisfaction from EMA smoking events) and lower levels of Stage 1 slope (satisfaction from EMA ecig relative to smoking events) lead to increased smoking rates

Stage 2 analysis - averaged regression results from 500 plausible value replications of Stage 1 random effects; controlling for baseline level of smoking dependency (`cigNDSSc`)

Variable	Estimate	Std Error	z-value	p-value
Intercept	2.36837	0.06845	34.60029	0.00000
cigNDSSc	0.62973	0.10599	5.94145	0.00000
Stage 1 Intercept	0.09163	0.07208	1.27135	0.20360
Stage 1 Slope	-0.28249	0.07397	-3.81908	0.00013
Stage 1 Scale	0.03892	0.07205	0.54023	0.58904
Residual Variance	1.10056	0.10174	10.81791	0.00000

$\Rightarrow$ higher levels of Stage 1 slope (satisfaction from EMA ecig relative to smoking events) lead to decreased smoking rates

Stage 2 analysis - N = 240 subjects - Average Daily ECig Rate

```
Descriptives
------------


 Dependent variable
                            mean          min          max      std dev
 ------------------------------------------------------------------
 sr_ECIGRate_mean         1.7402       0.0000       6.9725       1.1029


 Random Location and Scale EB mean estimates
                            mean          min          max      std dev
 ------------------------------------------------------------------
 Locat_1                  0.0000      -2.5639       1.6546       0.9706
 Locat_2                 -0.0000      -5.1979       3.4499       0.9170
 Scale                   -0.0000      -4.1966       1.7523       0.9481
```

Stage 2 analysis - averaged regression results from 500 plausible value replications of Stage 1 random effects

Variable	Estimate	Std Error	z-value	p-value
Intercept	1.74023	0.06938	25.08351	0.00000
Stage 1 Intercept	-0.14323	0.07282	-1.96676	0.04921
Stage 1 Slope	0.17843	0.07473	2.38754	0.01696
Stage 1 Scale	-0.12727	0.07290	-1.74573	0.08086
Residual.Variance	1.14158	0.10529	10.84252	0.00000

- lower levels of Stage 1 intercept (satisfaction from EMA smoking events) and higher levels of Stage 1 slope (satisfaction from EMA ecig relative to smoking events) lead to increased ECIG rates
- increased volatility of EMA reports leads to decreased ECIG rates

Stage 2 analysis - averaged regression results from 500 plausible value replications of Stage 1 random effects; controlling for baseline level of ECIG dependency (`EcigNDSSc`)

Variable	Estimate	Std Error	z-value	p-value
Intercept	1.74032	0.06676	26.06816	0.00000
ECigNDSSc	0.38619	0.08947	4.31648	0.00002
Stage 1 Intercept	-0.14264	0.06974	-2.04516	0.04084
Stage 1 Slope	0.14447	0.07246	1.99386	0.04617
Stage 1 Scale	-0.12460	0.07025	-1.77371	0.07611
Residual.Variance	1.05838	0.09746	10.85917	0.00000

- lower levels of Stage 1 intercept (satisfaction from EMA smoking events) and higher levels of Stage 1 slope (satisfaction from EMA ecig relative to smoking events) lead to increased ECIG rates
- increased volatility of EMA reports leads to decreased ECIG rates

Summary

- Joint shared parameter models are popular, but concern about time incompatibilities

- Two-stage approach can be more logically consistent

- Focus is on modeling Stage 1 intensive longitudinal data in terms of means (intercepts & slopes) and variance

- Stage 2 outcome is at subject level and predictors are Stage 1 random subject effects (and other subject variables)

- Freeware software program MixWILD automates the two-stage modeling, including random draws of plausible values of Stage 1 random effects: https://voices.uchicago.edu/hedeker/

Dzubur E, Ponnada A, Nordgren R, Yang CH, 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*:1403–1427..