Research Article

Statistics in Medicine

Received 22 August 2011, Accepted 13 March 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5393

A three-level mixed-effects location scale model with an application to ecological momentary assessment data

Xue Li[a*†] and Donald Hedeker[b]

In studies using ecological momentary assessment (EMA), or other intensive longitudinal data collection methods, interest frequently centers on changes in the variances, both within-subjects and between-subjects. For this, Hedeker *et al.* (Biometrics 2008; 64: 627–634) developed an extended two-level mixed-effects model that treats observations as being nested within subjects and allows covariates to influence both the within-subjects and between-subjects variance, beyond their influence on means. However, in EMA studies, subjects often provide many responses within and across days. To account for the possible systematic day-to-day variation, we developed a more flexible three-level mixed-effects location scale model that treats observations within days within subjects, and allows covariates to influence the variance at the subject, day, and observation level (over and above their usual effects on means) using a log-linear representation throughout. We provide details of a maximum likelihood solution and demonstrate how SAS PROC NLMIXED can be used to achieve maximum likelihood estimates in an alternative parameterization of our proposed three-level model. The accuracy of this approach using NLMIXED was verified by a series of simulation studies. Data from an adolescent mood study using EMA were analyzed to demonstrate this approach. The analyses clearly show the benefit of the proposed three-level model over the existing two-level approach. The proposed model has useful applications in many studies with three-level structures where interest centers on the joint modeling of the mean and variance structure. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: heteroscedasticity; location scale model; log-linear variance; multilevel data; variance modeling; maximum marginal likelihood

1. Introduction

For longitudinal data, mixed-effects regression models usually include random subject effects to account for the similarity among repeated measures for a given subject. The variance of the random subject effects, which represents between-subjects (BS) variation and the error variance, which represents within-subjects (WS) variation, are usually considered to be homogeneous across subject groups or levels of covariates. However, in reality, the homogeneous variance assumption, both within-subjects and between-subjects, can be violated, and the random subject effects can further be correlated with the error terms. Nonhomogeneous variance is often referred to as heteroscedasticity. By allowing for heteroscedasticity of within-subjects and between-subjects variation, the standard errors of the fixed-effects parameter estimates may be reduced, sometimes dramatically, and the precision of the estimation therefore increased.

Modern data collection procedures, such as ecological momentary assessments (EMA) [1, 2], experience sampling [3], and diary methods [4], have been developed to record the momentary events and experiences of subjects in daily life. These procedures yield relatively large BS and WS data, which allows the possibility to assess intraindividual variability and changes in the variances. The data from such designs are sometimes referred to as intensive longitudinal data [5].

[a] *VA Cooperative Studies Program Coordinating Center, Hines, IL 60141, USA*
[b] *School of Public Health, University of Illinois at Chicago, Chicago, IL 60612, USA*
**Correspondence to: Xue Li, Cooperative Studies Program Coordinating Center, Veteran Affairs, 5000 South 5th Ave, Bldg. 1 Section B, Room B203, Hines, IL 60141–30302, USA.*
† *E-mail: xue.li@va.gov*

Data from EMA usually have up to 30 or 40 observations per subject, and are inherently multilevel; for example, (level 1) occasions nested within (level 2) subjects, or more accurately, (level 1) occasions nested within (level 2) days, which are in turn nested within (level 3) subjects. Thus, linear mixed models are increasingly used for EMA data analysis [5, 6]. A particular interest in EMA studies is the modeling of BS and WS variances as a function of covariates, in addition to their effect on the overall mean levels. For this, Cleveland *et al*. [7, 8] first proposed a general class of models, mixed-effects location scale (LS) models, for WS variance modeling, which includes one or more random effects to characterize an individual's mean response (location), and an additional random (scale) effect in the error variance to characterize the variability around an individual's mean response. Following their work, two-level mixed-effects LS models have been described allowing for the effects of covariates on both the WS and BS variances [9, 10] but without including a random scale effect in the error variance. The recent work by Hedeker and others [11, 12] built upon this previous work by including a random scale effect in the error variance and also allowing for correlation between the random location and scale effects. However, an aspect that is ignored in all these two-level analyses of EMA data is the possibility of systematic day-to-day variation. As noted above, the observations are also nested within days (and subjects) and such day-to-day variation is simply treated as part of the WS variance in a two-level model. For example, a person's mood can vary from day to day, and within a day. Thus, mood can vary between subjects (some feel happy and some feel sad), within subjects but between days (some days are better and some days are worse for a given subject), and within subjects and days (mood can vary across the hours of the day for a given subject). A three-level model that treats occasions within days within subjects and separates between-day and within-day variation therefore represents a fuller examination for the analysis of EMA data. However, the existing methodology and software does not exist for such a general three-level model. Therefore, there is a need for developing methods for a three-level analysis with general variance modeling and random scale effects, and providing a convenient software program that is accessible via the major packages such as SAS.

In this article, we develop a three-level mixed-effects LS model with a three-level structure as: occasions (level 1) nested within days (level 2) nested within subjects (level 3). The proposed model is based on a conventional three-level mixed-effects regression model with a random intercept at each level, but also allows covariates to influence the variances at each level, using a log-linear representation throughout. The error variance is further allowed to vary across subjects above and beyond the contribution of covariates through a normally distributed random (scale) effect. Thus, the error variance follows a log-normal distribution. Furthermore, the random scale effect is allowed to be correlated with the random location effect. We demonstrate how the SAS (SAS Institute Inc., Cary, NC, USA) procedure PROC NLMIXED can be used to obtain maximum likelihood estimates of the proposed three-level model by reformulating the three-level model in a two-level formulation through a multivariate conditional likelihood approach. A syntax example is provided in the Appendix to facilitate this. The accuracy of this approach using NLMIXED was verified using a series of simulation studies. We also provide details of the maximum likelihood (ML) solution using a Fisher Scoring algorithm and Gauss–Hermite quadrature, so that researchers can produce their own programs using other software platforms. The proposed three-level mixed-effects LS model is illustrated using data from an EMA adolescent mood study, where interest is on determinants of the variation in the adolescents' moods.

2. Motivating adolescent mood study example

The data that motivated the development of the three-level mixed-effects LS model are from a study of mood among adolescents. Subjects were either 9th or 10th graders at baseline, 55.1% female, and self-reported on a screening questionnaire, 6–8 weeks prior to baseline, that they had smoked at least one cigarette in their life. The majority (57.6%) had smoked at least one cigarette in the past month at baseline. The study used a multimethod approach to assess adolescents at multiple time points in terms of self-report questionnaires, in-depth interviews, and week-long EMA sampling via hand-held palmtop computers. Here, we focus on the data from the baseline EMA collection. Adolescents carried hand-held computers for a seven-consecutive-day data collection period and were trained to respond to random prompts from the computers and also to self-initiate event recordings of smoking episodes. In this article, we focus on the responses from random prompts, which were date-stamped and time-stamped, and were initiated by the device approximately four times per day (range 1–8). Questions asked about location, activity, companionship, mood, and other behaviors. A total of 14,105 random prompts were obtained on 3642 days from 461 subjects with an approximate average of 30 prompts per subject (range = 7 to 71).

For the analyses reported, a three-level structure of random prompts (level 1) within measured days (level 2) within subjects (level 3) was considered. Some information from the self-initiated smoking events was used as covariates.

We considered two continuous outcomes: measures of the subject's negative affect (NA) and positive affect (PA) before the prompt signal. Both NA and PA consisted of the average of several mood items, each rated from 1 to 10 (with '10' representing very high levels of the attribute). NA consisted of five items to assess preprompt negative mood: I felt sad, I felt stressed, I felt angry, I felt frustrated, and I felt irritable; and PA consisted of five items assessing positive mood just before the prompt: I felt happy, I felt relaxed, I felt cheerful, I felt confident, and I felt accepted by others. Higher NA score reflects relatively poorer moods; whereas higher PA scores indicate relatively better mood. In this study, interest focuses on the degree to which covariates explain between-subject (BS), within-subject between-day (WS-BD), and within-subject within-day (WS-WD) variation in NA and PA, over and above their influence on the mean response. In particular, the effect of a subject's smoking level on both the mean response and variance heterogeneity were examined.

3. Three-level mixed-effects location scale model

To describe the model, the three-level data structure is defined as follows: Assume that there are $k = 1, \ldots, n_{ij}$ level 1 units that are nested within $j = 1, \ldots, n_i$ level 2 units that are in turn nested within $i = 1, \ldots, n$ level 3 units. The three-level mixed-effects model with random intercepts at both levels 2 and 3 can be written as

$$y_{ijk} = \boldsymbol{x}_{ijk}^T \boldsymbol{\beta} + \gamma_i + \upsilon_{ij} + \varepsilon_{ijk}. \tag{1}$$

In the adolescent mood study that was used as an example for motivation and illustration, occasions (level 1) were nested within days (level 2) and nested within subjects (level 3). The outcome y_{ijk} is the mood affect measurement, either negative affect or positive affect, of subject i on day j and on occasion k. The covariate vector $\boldsymbol{x}_{ijk}$(first element is one) includes levels 1, 2, and 3 explanatory variables, and $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients. The random subject effect γ_i indicates the mood affect influence of subject i, while the random day effect υ_{ij} represents the influence on mood of subject i on day j. These random effects are referred to as random location effects in the context of LS models. The population distributions of γ_i and υ_{ij} are assumed to be normal distributions with $N(0, \sigma_\gamma^2)$ and $N\left(0, \sigma_\upsilon^2\right)$, respectively. The errors ε_{ijk} are also assumed to be normal with $N(0, \sigma_e^2)$ and independent of the random effects at both level 2 and 3. In the adolescent mood example, σ_e^2 is the WS-WD variance, σ_υ^2 represents the WS-BS variance and σ_γ^2 is the BS variance. Because the level 3 subscript i is present for both n_i and n_{ij}, not all level 2 units are assumed to have the same number of level 1 units nested within, and not all level 3 units are assumed to have the same number of level 2 units nested within. In other words, there is no assumption of equal sample size at any level.

By allowing for heteroscedasticity of random effect (levels 2 and 3) variances and error variances at each level, we can further allow covariates to influence these variances. As such, we can utilize a log-linear representation, as has been described in the context of heteroscedastic regression models [13, 14], namely,

$$\log\left(\sigma_{\gamma_i}^2\right) = \boldsymbol{\pi}_i^T \boldsymbol{\lambda}, \quad \log\left(\sigma_{\upsilon_{ij}}^2\right) = \boldsymbol{u}_{ij}^T \boldsymbol{\alpha} \quad \text{and} \quad \log\left(\sigma_{\varepsilon_{ijk}}^2\right) = \boldsymbol{\varpi}_{ijk}^T \boldsymbol{\tau}. \tag{2}$$

The variances are subscripted by i, j, and k to indicate that their values change depending on the values of the covariates $\boldsymbol{\pi}_i$, $\boldsymbol{u}_{ij}$, and $\boldsymbol{\varpi}_{ijk}$ (and their coefficients), which include a (first) column of ones. The number of parameters associated with these variances does not vary with i or j or k. Thus, the variance of level 3 random effect $\sigma_{\gamma_i}^2$ equals $\exp(\lambda_0)$ when the level 3 covariates $\boldsymbol{\pi}_i$ equal $\boldsymbol{0}$, and is increased or decreased as a function of these covariates and their coefficients $\boldsymbol{\lambda}$. We chose a log function here to ensure that the variance would be positive. The variances of level 2 random effects and the error variance are modeled in the same way, except that both levels 2 and 3 covariates ($\boldsymbol{u}_{ij}$) may influence $\sigma_{\upsilon_{ij}}^2$; and covariates from all three levels ($\boldsymbol{\varpi}_{ijk}$) may influence $\sigma_{\varepsilon_{ijk}}^2$. The coefficients in $\boldsymbol{\lambda}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\tau}$ indicate the degree of influence on the variances $\sigma_{\gamma_i}^2, \sigma_{\upsilon_{ij}}^2$, and $\sigma_{\varepsilon_{ijk}}^2$, respectively, and the ordinary three-level random intercept model is obtained as a special case if $\boldsymbol{\lambda} = \boldsymbol{\alpha} = \boldsymbol{\tau} = \boldsymbol{0}$ for all covariates in $\boldsymbol{\pi}_i, \boldsymbol{u}_{ij}$, and $\boldsymbol{\varpi}_{ijk}$, excluding the reference variance λ_0, α_0, and τ_0.

The error variance $\sigma^2_{\varepsilon_{ijk}}$ can be further modeled to vary across individuals, above and beyond the contribution of covariates, namely,

$$\log\left(\sigma^2_{\varepsilon_{ijk}}\right) = \boldsymbol{\varpi}^T_{ijk}\boldsymbol{\tau} + \omega_i, \tag{3}$$

where the random level 3 scale effects ω_i have a normal distribution with zero mean and variance σ^2_ω. Thus, the variance $\sigma^2_{\varepsilon_{ijk}}$ is a random variable that has a log-normal distribution. The choice of the skewed and non-negative log-normal distribution for $\sigma^2_{\varepsilon_{ijk}}$ has been used in many diverse research areas for representing variances [15–19]. In this model, the random location effects γ_i and υ_{ij} indicate how level 3 units differ in terms of mean levels, while the random scale effects ω_i indicate how level 3 units differ in variation, beyond the effect of covariates. Thus, the model with both types of random effects (random location and random scale) is referred to as a mixed-effects LS model. The level 3 random location effects γ_i and random scale effects ω_i are further correlated with covariance parameter $\sigma_{\gamma\omega}$, and each is independent of level 2 random location effects υ_{ij}. This covariance parameter indicates the degree to which level 3 random location and scale effects are associated. In the adolescent mood study, it shows how a subject's mood affect mean is associated with their mood variation. Jointly, υ_{ij}, γ_i, and ω_i are assumed to have a multivariate normal distribution with zero mean and variance–covariance matrix as described in (4).

$$\begin{bmatrix} \upsilon_{ij} \\ \gamma_i \\ \omega_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\upsilon_{ij}} & 0 & 0 \\ 0 & \sigma^2_{\gamma_i} & \sigma_{\gamma\omega} \\ 0 & \sigma_{\gamma\omega} & \sigma^2_\omega \end{bmatrix}\right). \tag{4}$$

For estimation purposes, the random effects are usually expressed in standardized form (i.e., as multivariate standard normal) using the Cholesky factorization, namely

$$\begin{pmatrix} \upsilon_{ij} \\ \gamma_i \\ \omega_i \end{pmatrix} = \begin{bmatrix} \sigma_{\upsilon_{ij}} & 0 & 0 \\ 0 & \sigma_{\gamma_i} & 0 \\ 0 & \sigma_{\gamma\omega} / \sigma_{\gamma_i} & \sqrt{\sigma^2_\omega - \sigma^2_{\gamma\omega} / \sigma^2_{\gamma_i}} \end{bmatrix} \begin{bmatrix} \theta_{ij} \\ \theta_{1i} \\ \theta_{2i} \end{bmatrix}. \tag{5}$$

The model (1) can be rewritten as $y_{ijk} = \boldsymbol{x}^T_{ijk}\boldsymbol{\beta} + \sigma_{\gamma_i}\theta_{1i} + \sigma_{\upsilon_{ij}}\theta_{ij} + \sigma_{\varepsilon_{ijk}} e_{ijk}$, where e_{ijk} has a standard normal distribution, and $\sigma_{\varepsilon_{ijk}}$ is the standard deviation of ε_{ijk} given ω_i and is expressed as

$$\sigma_{\varepsilon_{ijk}} = \exp\left\{\frac{1}{2}\left(\boldsymbol{\varpi}^T_{ijk}\boldsymbol{\tau} + \sigma_{\gamma\omega} / \sigma_{\gamma_i} \cdot \theta_{1i} + \sqrt{\sigma^2_\omega - \sigma^2_{\gamma\omega} / \sigma^2_{\gamma_i}} \cdot \theta_{2i}\right)\right\}. \tag{6}$$

The random effects θ_{1i}, θ_{2i}, and θ_{ij} are pairwise independent, and each follows a univariate normal distribution with zero mean and unit variance, and each is independent of e_{ijk}. Given ω_i the error ε_{ijk} has a normal distribution with zero mean and variance $\sigma^2_{\varepsilon_{ijk}} = \exp(\boldsymbol{\varpi}^T_{ijk}\boldsymbol{\tau} + \omega_i)$. The marginal distribution of ε_{ijk}, however, is no longer normal and instead it has a complex form involving the product of a log-normal random variable $\sigma_{\varepsilon_{ijk}}$ with a standard normal random variable e_{ijk}. As such, the marginal distribution of y_{ijk} is also not normal and its marginal mean and variance are, respectively, $\boldsymbol{x}^T_{ijk}\boldsymbol{\beta}$ and $\exp\left(\boldsymbol{u}^T_{ij}\boldsymbol{\alpha}\right) + \exp\left(\boldsymbol{\pi}^T_i\boldsymbol{\lambda}\right) + \exp\left(\boldsymbol{\varpi}^T_{ijk}\boldsymbol{\tau} + 0.5\sigma^2_\omega\right)$.

4. Maximum marginal likelihood estimation

Parameters are estimated using a maximum marginal likelihood (MML) estimation method. The details about the estimation are described in Appendix A. Fisher's method of scoring can be used to provide the solution to the likelihood equations. For this, provisional estimates for the vector of parameters $\boldsymbol{\psi} = \left(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \sigma_{\gamma\omega}, \sigma^2_\omega\right)^T$ on iteration t are improved by

$$\boldsymbol{\psi}_{t+1} = \boldsymbol{\psi}_t - E\left(\frac{\partial^2 \log L}{\partial\boldsymbol{\psi}_t \partial\boldsymbol{\psi}^T_t}\right)^{-1} \frac{\partial L}{\partial\boldsymbol{\psi}_t}, \tag{7}$$

where the information matrix, or minus the expectation of the matrix of second derivatives, is

$$- E \left(\frac{\partial^2 \log L}{\partial \boldsymbol{\psi}_t \partial \boldsymbol{\psi}_t^T} \right) = \sum_{i=1}^{n} h^{-2}(y_i) \frac{\partial h(\boldsymbol{y}_i)}{\partial \boldsymbol{\psi}_t} \left(\frac{\partial h\,(\boldsymbol{y}_i)}{\partial \boldsymbol{\psi}_t} \right)^T . \tag{8}$$

The right-hand side is often referred to as the BHHH method because of Berndt *et al.* [20]. At convergence, the large-sample variance–covariance matrix of the parameter estimates is then obtained as the inverse of the information matrix in (8). In the MML solution, numerical integration is performed on the transformed $\boldsymbol{\theta}^*$ space, specifically Gauss–Hermite quadrature can be used to approximate the integrals by summations on Q quadrature points for each dimension of the integration. Further details on the quadrature approach are described in [21] and [22].

5. Computer implementation

The MML solution presented in Appendix A can be programmed using, say, FORTRAN (Intel Compliers) or C++. Alternatively, the SAS program PROC NLMIXED can also be used for MML estimation. In general, three-level models cannot be fit in PROC NLMIXED, which allows only a single level of random effects. However, the general statement in PROC NLMIXED allows one to write a multivariate conditional likelihood and hence can fit a three-level model. For this, the log-likelihood needs to be derived in a closed form and the dataset needs to be reshaped into a two-level form. In this regard, Gueorguieva [23] reformulated a three-level correlated probit model (level 1 repeated measures nested within level 2 fetus, which in turn are nested within level 3 litters) to a two-level form, and fit the reformulated two-level model in PROC NLMIXED. Her approach, however, only allows for two level 1 observations nested within each level 2 fetus. Alternatively, the SAS program developed here allows for unlimited level 1 observations nested within level 2 units, in turn nested within level 3 units. The details of this approach are described in Appendix B. A SAS syntax example for the proposed three-level mixed-effects LS model, and the first 13 observations from a hypothetical dataset used to illustrate the construction of the conditional log-likelihood, is provided in Appendix C. The starting values for parameters $\beta_0, \lambda_0, \alpha_0$, and τ_0 were obtained from a three-level random intercept model without any covariates at the mean level, and the random scale variance and covariance terms (σ_ω^2 and $\sigma_{\gamma\omega}$) were estimated from a two-level mixed-effects LS model without any covariates. All the remaining parameters were set to 0s. More details are provided in Appendix C.

6. Simulation study

A series of simulations using 1000 data sets, each with 11,200 observations (4 level 1 observations within each of the 7 level 2 units within each of the 400 level 3 units), were generated under the proposed three-level mixed-effects LS model with three covariates (either continuous or dichotomous), one at each level, to assess the accuracy and reliability of the proposed three-level model using PROC NLMIXED. We also compared its performance to a simpler three-level random intercept model. These covariates were specified to have effects on the variances and the mean. Therefore, the fixed-effects covariate vector x_{ijk} and the covariate vector $\boldsymbol{\varpi}_{ijk}$ in error variance include covariates from levels 1, 2, and 3 (i.e., X1, X2, and X3, respectively); the covariate vector $\boldsymbol{u}_{ij}$ in level 2 random effect variance includes level 2 (X2) and level 3 (X3) covariates; and the covariate vector $\boldsymbol{\pi}_i$ in level 3 random effect variance includes the level 3 covariate X3. Three levels (small, medium, large) of random scale variance $\left(\sigma_\omega^2\right)$ along with two opposite covariances ($\sigma_{\upsilon\omega}$) were further evaluated. Standardized biases (SBs), root mean square errors (RMSEs), 95% CI coverage probabilities, and average lengths of 95% CIs over the 1000 unique datasets were assessed to evaluate model performance as described by Demirtas [24].

We present in Table I the results from one set of simulations, which included three continuous covariates: X1 $\sim N$ $(\mu = 0.5, \sigma = 0.5)$, X2 $\sim N$ $(\mu = -0.2, \sigma = 1.2)$, and X3 $\sim N$ $(\mu = 0, \sigma = 0.7)$ at levels 1, 2, and 3, respectively. The simulation results reveal that the proposed three-level LS model using PROC NLMIXED recovers the assigned parameter values ($\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \sigma_{\upsilon\omega}$) adequately, as indicated by small biases and RMSEs, acceptable standardized bias, and close to 95% coverage. We obtained similar findings in all other scenarios that, for space, are not presented here. These additional results are available at http://www.uic.edu/classes/bstt/bstt513/pubs.html. For the variance of the random scale effect σ_ω^2, the average of the estimated values is close to the true value, but the coverage tends to decrease to around

Table I. Results from 1000 simulations under the three-level mixed-effects location scale model — three continuous covariates and small random scale effect and positive covariance.

		Three-level random intercept model							Three-level mixed-effects location scale model						
Parameters	True value	EST	SE*	Bias	SB	RMSE	95% COV	AW	EST	SE*	Bias	SB	RMSE	95% COV	AW
β_0 : (Intercept)	6.90	6.899	0.06	–0.0012	–2.16	0.06	95.9	0.23	6.899	0.06	–0.0011	–1.99	0.06	95.8	0.23
β_1 : (X1)	–0.40	–0.400	0.03	0.0002	0.80	0.03	95.4	0.11	–0.400	0.02	0.0003	1.13	0.02	94.5	0.10
β_2 : (X2)	0.20	0.200	0.01	0.0001	0.50	0.01	95.5	0.06	0.200	0.01	0.0001	0.84	0.01	95.8	0.06
β_3 : (X3)	0.60	0.603	0.08	0.0031	3.66	0.08	94.8	0.32	0.604	0.08	0.0037	4.41	0.08	94.3	0.32
λ_0 : (Intercept)	0.20	0.194	0.08	–0.0056	–7.34	0.08	96.2	0.30	0.186	0.08	–0.0136	–17.90	0.08	96.1	0.30
λ_1 : (X3)	–0.10	—	—	—	—	—	—	—	–0.100	0.11	0.0002	0.19	0.11	94.4	0.42
α_0 : (Intercept)	–1.20	–1.136	0.08	0.0641	77.06	0.10	82.3	0.29	–1.199	0.07	0.0011	1.62	0.07	95.2	0.28
α_1 : (X2)	–0.10	—	—	—	—	—	—	—	–0.102	0.06	–0.0019	–3.11	0.06	94.2	0.24
α_2 : (X3)	–0.40	—	—	—	—	—	—	—	–0.404	0.10	–0.0037	–3.68	0.10	94.4	0.38
τ_0 : (Intercept)	0.40	0.635	0.04	0.2351	644.84	0.24	0.0	0.06	0.398	0.04	–0.0015	–4.26	0.04	94.8	0.14
τ_1 : (X1)	0.10	—	—	—	—	—	—	—	0.099	0.03	–0.0007	–2.44	0.03	96.4	0.13
τ_2 : (X2)	–0.10	—	—	—	—	—	—	—	–0.100	0.01	–0.0002	–1.32	0.01	94.7	0.05
τ_3 : (X3)	–0.20	—	—	—	—	—	—	—	–0.200	0.04	–0.0004	–0.89	0.04	95.6	0.18
σ^2_ω	0.30	—	—	—	—	—	—	—	0.289	0.03	–0.0106	–38.36	0.03	91.0	0.11
$\sigma_{\upsilon\omega}$	0.15	—	—	—	—	—	—	—	0.151	0.04	0.0011	2.88	0.04	94.2	0.14
–2LogL				41338.5							40326.8				
# of converged solutions					999							975			

∗ EST: estimate; SE: standard error; SB: standardized bias; RMSE: root mean square error; COV: coverage probability; AW: average width.

86% when the size of the random scale effect is large. In general, as is well known, using the Wald test for a variance parameter is not ideal. Except for somewhat minor underestimation of the random scale variance parameter, the inferences and conclusions for all other parameters are fairly good using PROC NLMIXED.

The results from the traditional three-level random intercept model, which ignores the random scale effects and covariate effects on variances, indicate good recovery of the mean effects ($\boldsymbol{\beta}$). However, this model yields badly biased estimates of the variance parameters. As is demonstrated in Table I, when the level 2 covariate X2 and the level 3 covariate X3 are ignored in modeling the variance $\sigma^2_{\upsilon_{ij}}$, the standardized bias of $\hat{\alpha}_0$ can be as large as 77, and the coverage probability for this parameter drops to around 80%. Also, ignoring the random scale effect and the covariate effects (X1, X2, and X3) on the error variance ($\sigma^2_{\varepsilon_{ijk}}$) modeling, the standardized bias of $\hat{\tau}_0$ can be as high as 800 and the 95% coverage probabilities in all scenarios drop to 0%. The huge positive standardized biases indicate that the estimate of τ_0 on average falls about five standard deviations for small random scale ($\sigma^2_\omega = 0.30$), and about eight standard deviations for large random scale ($\sigma^2_\omega = 1.20$), above the true parameter. Additionally, the bias, standardized bias, and RMSE tend to increase when the random scale variance increases. Our simulations only consider one covariate at each level, thus if the number of covariates at each level increases, the estimated intercepts of level 3, 2 random effects variances and error variance ($\hat{\lambda}_0, \hat{\alpha}_0$, and $\hat{\tau}_0$) could be even more biased (if these covariates are ignored in the variance modeling).

7. Application to adolescent mood data

Data from the EMA Adolescent Mood Study are used to illustrate the application of the proposed three-level mixed-effects LS model. Here, we focus on the degree to which covariates might explain variation at the subject-level, day-level, and prompt-level in NA and PA, over and above their influence on the mean response. Subject-level (level 3) covariates include **Smoker** (defined as the presence of at least one smoking event during the EMA baseline data collection period, $1 =$ yes or $0 =$ no), **PropSmk** (a proportion that indicates the level of smoking and is defined as the number of smoking events over the total number of random prompts and smoking events), **Male** ($1 =$ male or $0 =$ female), **Grade10** ($1 =$ 10th or $0 =$ 9th grade), **NovSeekC** (a measure of novelty seeking), and **NegMoodRegC** (a measure of negative mood regulation). Among these covariates, **NovSeekC** and **NegMoodRegC** are grand mean centered. Day-level (level 2) covariates include **WeekEnd** ($0 =$ weekday indicating Monday to Friday or $1 =$ weekend indicating Saturday and Sunday). For prompt-level (level 1) covariates, we considered whether the subject was alone or accompanied by others ($0 =$ not alone or $1 =$ alone) at the time of the random prompt. For this variable, we created both BS and WS-WD versions, **AloneBS** and **AloneWS**, as described by Neuhaus and Kalbfeisch [25], namely the decomposition $X_{ijk} = \bar{X}_{i..} + (X_{ijk} - \bar{X}_{i..})$. Here, **AloneBS** $= \bar{X}_{i..}$ equals the proportion of random prompts in which a subject was alone, and **AloneWS** $= X_{ijk} - \bar{X}_{i..}$ is the prompt-specific deviation from the proportion. Note that **AloneBS** is a subject-level covariate and **AloneWS** is a prompt-level covariate.

Results are given in Table II for NA and Table III for PA. For comparison purposes, estimates of a two-level mixed-effects LS model and a three-level random intercept model are also listed. The first column lists the estimates ($\hat{\boldsymbol{\beta}}$) and standard errors for the random-intercept model; the second to fourth columns, respectively, list the estimates ($\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\tau}}$) and standard errors for the two-level LS model; and the fifth to eight columns, respectively, list the estimates ($\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\tau}}$) and standard errors for the proposed three-level LS model. The variance estimates ($\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\tau}}$) are on the natural log scale and correspond to the regression parameter estimates for the BS, WS-BD, and WS-WD variances, respectively.

As can be seen from Tables II and III, for both NA and PA, the random scale effects (scale variance and location scale covariance) in the two-level and three-level LS models are highly significant ($p < 0.001$). Both AIC and BIC favor the three-level LS model relative to the two-level LS model, which in turn is favored relative to the three-level random intercept model. This provides clear evidence that the error variance varies across individuals, above and beyond the contribution of the many covariates, and that the three-level LS model outperforms the two-level LS model and both are better than the random intercept model. Also, in terms of the LS covariance, for NA the positive covariance estimate shows that subjects with higher NA mean (poorer mood) fluctuate more in NA, while for PA the negative covariance estimate indicates that subjects with higher PA mean (better mood) vary less across prompts in their PA responses.

Table II. Mixed-effects models of negative affect, $n = 461$ and $\sum_{ij} n_{ij} = 3462$ and $\sum_{ijk} n_{ijk} = 14150$, maximum likelihood estimates (standard errors).

		Mixed-effects location scale model						
	Random intercept	Two-level			Three-level			
Parameters	Mean (β)	Mean (β)	BS (λ)	WS (τ)	Mean (β)	BS (λ)	WS-BD (α)	WS-WD (τ)
Intercept	3.096***	3.094***	0.281	0.637***	3.128***	0.227	–0.660**	0.361*
	(0.200)	(0.191)	(0.215)	(0.138)	(0.191)	(0.227)	(0.206)	(0.145)
Smoker	0.367*	0.399*	0.187	0.372***	0.395*	0.138	0.572***	0.348**
	(0.162)	(0.162)	(0.156)	(0.112)	(0.160)	(0.167)	(0.144)	(0.117)
PropSmk	–1.060	–1.223	–0.427	–1.266*	–1.262	–0.469	–2.413**	–1.163*
	(0.766)	(0.750)	(0.752)	(0.530)	(0.731)	(0.812)	(0.785)	(0.561)
NovSeekC	0.183	0.184	–0.154	0.221**	0.188	–0.173	0.259**	0.203**
	(0.098)	(0.097)	(0.087)	(0.068)	(0.097)	(0.094)	(0.099)	(0.071)
NegMoodRegC	–0.784***	–0.767***	–0.239*	–0.274***	–0.759***	–0.234*	–0.730***	–0.194**
	(0.096)	(0.095)	(0.096)	(0.066)	(0.094)	(0.103)	(0.095)	(0.070)
Male	–0.408**	–0.352**	–0.224	–0.361***	–0.349**	–0.264*	–0.172	–0.388***
	(0.135)	(0.130)	(0.127)	(0.093)	(0.129)	(0.133)	(0.136)	(0.098)
Grade10	0.091	0.105	0.027	–0.069	0.096	0.013	0.059	–0.082
	(0.128)	(0.125)	(0.122)	(0.088)	(0.124)	(0.129)	(0.126)	(0.093)
AloneBS	0.926**	0.771*	0.512	0.357	0.776*	0.634	–0.525	0.531*
	(0.333)	(0.323)	(0.311)	(0.230)	(0.321)	(0.337)	(0.319)	(0.242)
AloneWS	0.347***	0.166***	—	0.050	0.165***	—	—	0.042
	(0.030)	(0.021)	—	(0.029)	(0.021)	—	—	(0.032)
WeekEnd	–0.228***	–0.102***	—	0.029	–0.202***	—	0.048	–0.006
	(0.045)	(0.020)	—	(0.032)	(0.036)	—	(0.146)	(0.034)
WS variance of scale σ^2_ω			0.801***(0.058)				0.872***(0.064)	
Covariance $\sigma_{\upsilon\omega}$			0.520***(0.061)				0.512***(0.062)	
–2logL	55058		52814				52086	
AIC	55084		52874				52164	
BIC	55138		52998				52326	

*** $p < 0.001$.
** $p < 0.01$.
* $p < 0.05$.

Table III. Mixed-effects models of Positive Affect, n= 461 and $\sum_{ij} n_{ij}$=3462 and $\sum_{ijk} n_{ijk}$=14150, maximum likelihood estimates (standard errors).

		Mixed-effects location scale model						
	Random intercept	Two-level			Three-level			
Parameters	Mean (β)	Mean (β)	BS (λ)	WS (τ)	Mean (β)	BS (λ)	WS-BD (α)	WS-WD (τ)
Intercept	7.443***	7.389***	–0.268	0.608***	7.385***	–0.284	–1.123***	0.421***
	(0.164)	(0.156)	(0.192)	(0.108)	(0.156)	(0.204)	(0.240)	(0.115)
Smoker	–0.133	–0.087	–0.068	0.144	–0.097	–0.041	0.636***	0.083
	(0.133)	(0.128)	(0.169)	(0.087)	(0.128)	(0.177)	(0.167)	(0.092)
PropSmk	–0.267	–0.261	0.667	–0.574	–0.254	0.709	–2.753**	–0.322
	(0.628)	(0.624)	(0.875)	(0.415)	(0.628)	(0.914)	(0.906)	(0.439)
NovSeekC	0.103	0.054	–0.256**	0.131*	0.054	–0.282**	0.244*	0.113*
	(0.080)	(0.080)	(0.094)	(0.053)	(0.081)	(0.098)	(0.111)	(0.056)
NegMoodRegC	0.589***	0.589***	–0.092	–0.163**	0.589***	–0.067	–0.288**	–0.143**
	(0.079)	(0.076)	(0.098)	(0.052)	(0.077)	(0.102)	(0.110)	(0.055)
Male	0.228*	0.180	–0.146	–0.230**	0.182	–0.176	–0.007	–0.260***
	(0.111)	(0.106)	(0.130)	(0.073)	(0.106)	(0.135)	(0.148)	(0.077)
Grade10	0.036	–0.012	–0.290*	–0.146*	–0.006	–0.289*	–0.181	–0.148*
	(0.105)	(0.102)	(0.123)	(0.069)	(0.102)	(0.128)	(0.142)	(0.073)
AloneBS	–1.485***	–1.311***	1.145***	0.304	–1.314***	1.096**	–0.269	0.401*
	(0.273)	(0.267)	(0.319)	(0.180)	(0.267)	(0.338)	(0.374)	(0.191)
AloneWS	–0.490***	–0.369***	—	0.053	–0.353***	—	—	0.041
	(0.027)	(0.023)	—	(0.028)	(0.023)	—	—	(0.031)
WeekEnd	0.209***	0.198***	—	–0.073*	0.203***	—	–0.162	–0.055
	(0.037)	(0.023)	—	(0.030)	(0.031)	—	(0.177)	(0.033)
WS variance of scale σ^2_ω			0.458*** (0.035)				0.506*** (0.039)	
Covariance $\sigma_{\upsilon\omega}$			–0.303*** (0.040)				–0.315*** (0.042)	
–2logL	52093		50560				50077	
AIC	52119		50620				50155	
BIC	52172		50744				50317	

$^{***}p < 0.001$.
$^{**}p < 0.01$.
$^{*}p < 0.0$.

In comparing the mean effects among the three models, some differences emerge between the random-intercept and the two mixed-effects LS models. In general, the former model yields a few more significant results than the latter. Specifically, for NA in Table II, the significance level of **AloneBS** in the random-intercept model (p-value < 0.01) is diminished in the two mixed-effects LS models as noted by the increasing p-values (< 0.05). Similarly, in Table III, the gender effect (**Male**) that significantly increases the PA mean in the random-intercept model is no longer significant (p-values > 0.05) in the two LS models. However, the differences between the two-level and three-level mixed-effects LS models are not obvious in the sense that the mean estimates from the two LS models are similar in magnitude and significance levels, and the standard errors of the parameter estimates are close. Thus, it would appear that if the main interest centers on changes in the mean, the two-level model works fairly well compared with the three-level model, and both are superior to the over-simplified random intercept model.

In comparing the two-level and three-level LS models, it is observed that most of the differences in variance modeling come from the parameter estimates associated with the error variance (i.e., WS variance in the two-level model and WS-WD variance in the three-level model). These differences arise from the separation of WS-BD and WS-WD variation in the three-level LS model, which are simply treated as error variance (i.e., WS variance) in the two-level model. As can be seen from Table II, for the NA outcome, the variables **Smoker**, **PropSmk**, and **NegMoodRegC** that significantly affect WS variance in the two-level model, have greater influence (i.e., smaller p-values) on persons' between-day variation (i.e., WS-BD variance), and less influence (i.e., larger p-values) on persons' within-day variation (i.e., WS-WD variance) in the three-level model. The variable **Male**, which has a significant negative effect on WS variance in the two-level model, only shows a significant negative effect on within-day variation but not on between-day variation in the three-level model. The nonsignificant variable **AloneBS** on WS variance in the two-level model becomes significant in terms of within-day but not between-day variance in the three-level model. The variable **NovSeekC**, which significantly increases WS variance in the two-level model, has the same significance level on both within-day and between-day variations in the three-level model.

Turning to the PA outcome, although currently smoking (**Smoker**) and the level of smoking (**PropSmk**) do not significantly affect WS mood variation in the two-level model, they significantly (p-values < 0.01) affect between-day variation, but not within-day variation in the three-level model. The variables **NovSeekC** and **NegMoodRegC** that either significantly increase or decrease the WS variance in the two-level model have equal influence (similar significance level) on both between-day and within-day variance in the three-level model. The variables **Male** and **Grade 10** that have significant negative effect on WS variation in the two-level model only show significant negative effect on within-day variation but not on between-day variation. The nonsignificant variable **AloneBS** on the WS variance in the two-level model becomes significant on within-day variation but not on between-day variation in the three-level model; conversely, the significant **WeekEnd** indicator on WS variance in the two-level model is no longer significant in the three-level model for both between-day and within-day variations. Thus, a three-level LS model provides a fuller examination of WS variation; we can more precisely assess where the within-subject variation occurs: either between-days, within-days, or both.

For a regular mixed-effects three-level model, we have several different kinds of ICCs that are of potential interest as described by Snijders and Bosker [26]. For the proposed three-level mixed-effects LS model, ICC estimates can also be obtained. The ICC at level 3 represents the proportion of total unexplained variation that is at the subject level and is also the correlation for two observations from the same individual on different days and is denoted as $\sigma^2_{\gamma_i} \,/\, \sqrt{Var(y_{ijk})Var(y_{ij'k'})}$. The ICC at levels 2 and 3 represents the proportion of total unexplained variation that is at both subject-level and day-level, and is also the correlation for two observations from the same individual on the same day. This ICC is denoted as $\left(\sigma^2_{\gamma_i} + \sigma^2_{\upsilon_{ij}}\right) / \sqrt{Var(y_{ijk})Var\left(y_{ijk'}\right)}$. Note that the ICC at either the subject-level or both subject-level and day-level can vary as a function of subject-level, day-level, and occasion-level covariates. Because the model allows the three variance components to vary as a function of covariates, here for simplicity, we report averaged variances at each level. Using these averaged variances, the ICC at the subject level (level 3) is estimated to be 0.337 for NA and 0.324 for PA, while the ICC at the day and subject level (levels 2 and 3) is estimated to be 0.443 for NA and 0.412 for PA. Therefore, of the total (unexplained) variance for NA, 33.7% is at the subject level, while 44.3% is at the day and subject level. Obviously, in terms of NA, there is more difference between subjects than within subjects (and across days), although the latter is not negligible. Similar conclusions apply to the ICCs for PA.

Statistics in Medicine

To summarize our findings, in terms of mean response, it is noted that several variables significantly increase (**Smoker**, **AloneBS**, **AloneWS**), and decrease (**NegMoodRegC**, **Male**, **WeekEnd**) the mean level of NA. Thus, being a smoker and a loner (i.e., higher on **AloneBS**) and being alone (i.e., higher on **AloneWS**) increase NA. Conversely, being a male and having better negative mood regulation (i.e., higher on **NegMoodRegC**) decrease NA. NA is also higher on weekdays and lower at weekends. Turning to PA, we found similar effects to those for NA (but in the opposite direction). Namely, the variables (**NegMoodRegC**, **WeekEnd**) significantly increase, whereas the variables (**AloneBS**, **AloneWS**) significantly decrease this mean. Thus, being a loner and being alone lower PA. Conversely, having better negative mood regulation increases PA. Also, PA is increased for weekends and decreased on weekdays. Turning to the covariate effects on the variances, in terms of BS variation, males and those with better negative mood regulation have less subject to subject NA variation and behave more homogeneously. Also, novelty seekers and 10th graders exhibit less PA variation, while loners have more PA variation. In terms of within-subjects day-to-day variation, all of the results observed for NA are also seen for PA. Namely, smokers and novelty seekers are more varied and less homogeneous from day to day, whereas negative mood regulators (i.e., higher on **NegMoodRegC**) and smokers with higher smoking levels (i.e., higher on **PropSmk**) vary less and behave more homogeneously in both NA and PA from day to day. As to WS-WD variation, several variables significantly increase this variance (**Smoker**, **NovSeekC**, **AloneBS**), whereas others significantly diminish this variance (**PropSmk**, **NegMoodRegC**, **Male**) for NA. Thus, for NA the WS-WD data are more varied from smokers, novelty seekers, and loners, and less varied from males, negative mood regulators, and smokers with higher smoking levels. Similarly, results were observed for PA, namely, the variables **NovSeekC** and **AloneBS** significantly increase this variance, while **NegMoodRegC** and **Male** significantly decrease this variance. In addition, 10th graders decrease WS-WD PA variance.

8. Discussion

In this article we have extended an existing two-level mixed-effects LS model proposed by Hedeker *et al.* [12] to three levels by adding an intermediate day level into the two-level structure (occasions nested within subjects) to account for within-subjects day-to-day mood variation in EMA data. The proposed three-level mixed-effects LS model therefore is based on a conventional three-level random intercept model, but allows covariates to influence the variances at the subject, day, and occasion levels. This model can examine the degree to which subjects are heterogeneous in terms of their mood variation by further including a subject-level random scale effect on the WS-WD variance. Our examples with NA and PA clearly show that subjects experience systematic mood variation from day to day and within days. In this article we also detailed how maximum likelihood estimation can be carried out using existing software (SAS PROC NLMIXED).

The methods developed in this article can easily generalize to a variety of EMA studies in smoking and cancer-relevant research areas, such as studying relapse among adolescent smokers [27], examining the urge to smoke [28] in the former, and the assessment of pain and symptoms, and diet and exercise in the latter. Because EMA studies typically involve many measurements obtained from subjects both within and across days, the three-level mixed-effects LS model would seem to be a useful tool for analysis of EMA data. Additionally, although the proposed model was developed for the analysis of an EMA dataset, the model can also be applied to other types of studies with three-level structures when the study interest involves covariate effects on the variances and the overall mean.

In this article, single random effects at each of the subject-level and day-levels were considered, but this could be generalized, for example, to allow random intercepts and trends. Specifically, we can generalize Equation (1) to allow covariates to influence multiple BS variance parameters (i.e., intercept and slope variances), multiple WS-BD variance parameters (i.e., intercept and slope variances), and the WS-WD variance, using a log-linear representation.

Modern data collection procedures, such as EMA, usually provide a fair amount of data, and so give rise to the opportunity for modeling of variances as a function of covariates. One might wonder about how many subjects and observations within subject data are necessary for estimation and variance modeling purposes. For random coefficient models, Longford [29] noted the difficulty with providing general guidelines about the degree of complexity, for the variation part of a model that a given dataset could support. This would also seem to be true here. Simulations with small sample sizes (e.g., 20 subjects with 5 observations each), gives the general impression that the main concern is that the algorithm does not often converge, but instead has estimation difficulties of one sort or another, in small sample situations.

The current work focuses on continuous outcomes. Further work could extend our three-level model to other types of outcomes such as binary or ordinal. For such outcomes, two-level mixed LS models for ordinal data are described by Hedeker *et al.* [11, 30], while Gibbons and Hedeker [21] and Raman and Hedeker [22] describe (random intercept) three-level mixed models for dichotomous and ordinal outcomes, respectively. Therefore, a future aim is to extend these approaches to develop a three-level mixed LS model for such categorical outcomes.

Our work here only considers a frequentist approach. Alternatively one could use a Bayesian approach, via MCMC, for estimation of the hierarchical variance components. A recent article describing a two-level model using a Bayesian approach is by Myles *et al*. [31].

Finally, it should be mentioned that affect levels, and variability in affect levels, could be influenced by other variables than those presented in the example of this paper. We included variables that were deemed to be 'good candidates', but we have not been exhaustive in our selection of covariates. Essentially, we feel that our example provides a reasonable approach for illustrating our statistical model, but certainly more work can be conducted to provide a more comprehensive modeling of mood variation. Also, there are other methodological approaches for modeling variance, for example work on linear oscillators [32, 33], typology clustering [34], and generalized mixed models [35]. Future work could compare the relative merits of these approaches.

Appendix A: Maximum marginal likelihood estimation

Let $\boldsymbol{y}_i$ be a vector of the responses from level 3 unit i on n_i level 2 units with n_{ij} level 1 units on each level 2 unit. The marginal density of $\boldsymbol{y}_i$ in the population can be expressed as the following integral of the conditional likelihood, $L(\cdot)$, weighted by the density $g\,(\cdot)$

$$h(\boldsymbol{y}_i) = \int_{\boldsymbol{\theta}_i^\dagger} L\left(\boldsymbol{y}_i | \boldsymbol{\theta}_i^\dagger\right) g\left(\boldsymbol{\theta}_i^\dagger\right) d\left(\boldsymbol{\theta}_i^\dagger\right), \tag{A1}$$

where $g\left(\boldsymbol{\theta}_i^\dagger\right)$ represents the distribution of the random effects $\boldsymbol{\theta}_i^\dagger = \left(\theta_{i1}, \ldots, \theta_{in_i}, \boldsymbol{\theta}_i\right)^T$ for unit i in the population. Here $\left(\theta_{i1}, \ldots, \theta_{in_i}\right)^T$ are the level 2 random effects with an n_i-dimensional standard multivariate normal distribution and $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2})$ are the level 3 random effects with a standard bivariate normal distribution. Jointly, $\boldsymbol{\theta}_i^\dagger$ has a $n_i + 2$ dimensional standard multivariate normal distribution. The subscript i only means that the dimension of $\boldsymbol{\theta}^\dagger$ changes with level 3 unit i. Assuming independence of the response vector $\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in_i}$ for unit i, conditional on the level 3 random effects $\boldsymbol{\theta}_i$, the marginal density becomes

$$h\left(\boldsymbol{y}_i\right) = \int_{\boldsymbol{\theta}_i} \left\{\prod_{j=1}^{n_i} \int_{\theta_{ij}} L_{ij}\left(\boldsymbol{\theta}_i^*\right) g\left(\theta_{ij}\right) \mathrm{d}\theta_{ij}\right\} g\left(\boldsymbol{\theta}_i\right) d\boldsymbol{\theta}_i, \tag{A2}$$

where θ_{ij} is a univariate standard normal random variable and represents the level 2 random effects for $j = 1, \ldots, n_i$. Further assuming independence of responses from n_{ij} observations for level 3 unit i on level 2 unit j conditional on its level 2 and 3 random effects θ_{ij} and $\boldsymbol{\theta}_i$, the conditional likelihood of $\boldsymbol{y}_{ij}$, that is, $L_{ij}\left(\boldsymbol{\theta}_i^*\right) = L(\boldsymbol{y}_{ij} | \theta_{ij}, \boldsymbol{\theta}_i)$, can be expressed as $\prod_{k=1}^{n_{ij}} f\left(y_{ijk} | \theta_{ij}, \boldsymbol{\theta}_i\right)$. Thus, with the independence assumption, the integration dimensionality is significantly reduced from original n_i+2 to 3. The marginal density of $\boldsymbol{y}_i$ can then be expressed as

$$h(\boldsymbol{y}_i) = \int_{\boldsymbol{\theta}_i} \left\{\prod_{j=1}^{n_i} \int_{\theta_{ij}} \prod_{k=1}^{n_{ij}} f\left(y_{ijk} \mid \theta_{ij}, \boldsymbol{\theta}_i\right) g(\theta_{ij}) \mathrm{d}\theta_{ij}\right\} g(\boldsymbol{\theta}_i) \mathrm{d}\boldsymbol{\theta}_i \tag{A3}$$

For parameter estimation, we differentiate the marginal log-likelihood for the patterns from the n level 3 units, $\log L\,(\boldsymbol{\psi}) = \sum_{i=1}^{n} \log h(\boldsymbol{y}_i)$, where $\boldsymbol{\psi} = \left(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \sigma_{\gamma\omega}, \sigma_{\omega}^2\right)^{\mathrm{T}}$ is the parameter vector to be estimated. Let $\boldsymbol{\eta}$ be a parameter vector or scalar of $\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\lambda}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\tau}^{\mathrm{T}}, \sigma_{\gamma\omega}$, and σ_{ω}^2, then we obtain

$$\frac{\partial \log L}{\partial \boldsymbol{\eta}} = \sum_{i=1}^{n} h^{-1}\left(y_i\right) \frac{\partial h(y_i)}{\partial \boldsymbol{\eta}}, \tag{A4}$$

where

$$\frac{\partial h(\boldsymbol{y}_i)}{\partial \boldsymbol{\eta}} = \int_{\theta_i} e_i \left\{ \sum_{j=1}^{n_i} e_{ij}^{-1} \int_{\theta_{ij}} \sum_{k=1}^{n_{ij}} L_{ij}\left(\boldsymbol{\theta}_i^*\right) \frac{\partial z_{ijk}}{\partial \boldsymbol{\eta}} g\left(\theta_{ij}\right) \mathrm{d}\theta_{ij} \right\} g\left(\boldsymbol{\theta}_i\right) \mathrm{d}\boldsymbol{\theta}_i. \tag{A5}$$

In (A5), e_{ij}, e_i, and z_{ijk} are defined as $e_{ij} = h\left(\boldsymbol{y}_{ij} \middle| \boldsymbol{\theta}_i\right) = \int_{\theta_{ij}} L_{ij}(\boldsymbol{\theta}_i^*) g\left(\theta_{ij}\right) \mathrm{d}\theta_{ij}, e_i = h\left(\boldsymbol{y}_i \middle| \boldsymbol{\theta}_i\right) = \prod_{j=1}^{n_i} e_{ij}$, and $z_{ijk} = \log f\left(y_{ijk} \middle| \theta_{ij}, \boldsymbol{\theta}_i\right) = -\log\sqrt{2\pi} - 0.5\log\sigma^2_{\varepsilon_{ijk}} - 0.5\varepsilon^2_{ijk}\sigma^{-2}_{\varepsilon_{ijk}}$, respectively. Furthermore, the first derivative of z_{ijk} with respective to $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, $\boldsymbol{\alpha}$,$\boldsymbol{\tau}$, $\boldsymbol{\sigma}_{\gamma\omega}$, and σ^2_ω are respectively,

$$\frac{\partial z_{ijk}}{\partial \boldsymbol{\beta}} = \frac{\varepsilon_{ijk}}{\sigma^2_{\varepsilon_{ijk}}} \boldsymbol{x}_{ijk}, \frac{\partial z_{ijk}}{\partial \boldsymbol{\lambda}} = \frac{1}{4} s_{2i} \boldsymbol{\pi}_i \left(\frac{\varepsilon^2_{ijk}}{\sigma^2_{\varepsilon_{ijk}}} - 1\right)\left(\frac{s_{2i}}{s_{3i}}\theta_{2i} - \theta_{1i}\right) + \frac{s_{1i}\varepsilon_{ijk}\boldsymbol{\pi}_i}{2\boldsymbol{\sigma}^2_{\varepsilon_{ijk}}}\theta_{1i},$$
$$\frac{\partial z_{ijk}}{\partial \boldsymbol{\alpha}} = \frac{s_{ij}\varepsilon_{ijk}\boldsymbol{u}_{ij}}{2\sigma^2_{\varepsilon_{ijk}}}\theta_{ij}, \frac{\partial z_{ijk}}{\partial \boldsymbol{\tau}} = \frac{1}{2}\left(\frac{\varepsilon^2_{ijk}}{\sigma^2_{\varepsilon_{ijk}}} - 1\right)\boldsymbol{\varpi}_{ijk},$$
$$\frac{\partial z_{ijk}}{\partial \sigma_{\gamma\omega}} = \frac{1}{2s_{1i}}\left(\frac{\varepsilon^2_{ijk}}{\sigma^2_{\varepsilon_{ijk}}} - 1\right)\left(\theta_{1i} - \frac{s_{2i}}{s_{3i}}\theta_{2i}\right), \frac{\partial z_{ijk}}{\partial \sigma^2_\omega} = \frac{1}{4s_{3i}}\left(\frac{\varepsilon^2_{ijk}}{\sigma^2_{\varepsilon_{ijk}}} - 1\right)\theta_{2i}.$$

Here, $\varepsilon_{ijk} = y_{ijk} - \boldsymbol{x}^{\mathrm{T}}_{ijk}\boldsymbol{\beta} - s_{ij}\theta_{ij} - s_{1i}\theta_{1i}, \sigma_{\varepsilon_{ijk}} = \exp\left\{\left(\boldsymbol{\varpi}^{\mathrm{T}}_{ijk}\boldsymbol{\tau} + \sigma_{\gamma\omega} / s_{1i} \cdot \theta_{1i} + s_{3i} \cdot \theta_{2i}\right) / 2\right\}$, $s_{ij} = \exp\left(\boldsymbol{u}^{\mathrm{T}}_{ij}\boldsymbol{\alpha} / 2\right)$, $s_{1i} = \exp\left(\boldsymbol{\tau}^{\mathrm{T}}_i\boldsymbol{\lambda}/2\right)$, $s_{2i} = \sigma_{\gamma\omega} / \exp\left(\boldsymbol{\pi}^T_i\boldsymbol{\lambda}/2\right)$, and $s_{3i} = \sqrt{\sigma^2_\omega - \sigma^2_{\gamma\omega}/\exp\left(\boldsymbol{\pi}^{\mathrm{T}}_i\boldsymbol{\lambda}\right)}$.

Appendix B: Computer implementation

Let $m_{ijk} = \upsilon_{ij} + \varepsilon_{ijk}$, the three-level mixed-effects LS model in Section 3 and Equation (1) can be rewritten as a two-level form as follows:

$$y_{ijk} = \boldsymbol{x}^{\mathrm{T}}_{ijk}\boldsymbol{\beta} + \gamma_i + m_{ijk}, \tag{B1}$$

where m_{ijk} is viewed as the new error term in the two-level form and $m_{ijk}\middle|\omega_i \sim N\left(0, \sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ijk}}\right)$. The column vector $\boldsymbol{y}_{ij}$ containing n_{ij} responses $(y_{ij1}, \ldots, y_{ijn_{ij}})^T$ from unit i on level 2 unit j is expressed as $\boldsymbol{y}_{ij} = \boldsymbol{x}_{ij}\beta + \mathbf{1}_{ij}\gamma_i + \boldsymbol{m}_{ij}$. Given ω_i, $\boldsymbol{m}_{ij} = (m_{ij1}, \ldots, m_{ijn_{ij}})^T$ has a multivariate normal distribution $N\left(\mathbf{0}, \sigma^2_{\upsilon_{ij}}\boldsymbol{J}_{ij} + \boldsymbol{D}_{ij}\right)$, where $\boldsymbol{J}_{ij}$ is a $n_{ij} \times n_{ij}$ square matrix with all elements being ones and $\boldsymbol{D}_{ij}$ is an $n_{ij} \times n_{ij}$ diagonal matrix with the kth diagonal element being $\sigma^2_{\varepsilon_{ijk}} = \exp\left(\boldsymbol{\omega}^{\mathrm{T}}_{ijk}\boldsymbol{\tau} + \omega_i\right)$ for $k = 1, \ldots, n_{ij}$. Following this, $\boldsymbol{y}_{ij}$ given γ_i and ω_i is also normal with $N\left(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \mathbf{1}_{ij}\gamma_i, \sigma^2_{\upsilon_{ij}}\boldsymbol{J}_{ij} + \boldsymbol{D}_{ij}\right)$. We assume that the n_i responses $(y_{i1}, \ldots, y_{in_i})^T$ from unit i are independent given level 3 random effects (ω_i and γ_i). By factoring the multivariate likelihood $f\left(\boldsymbol{y}_{ij}\middle|\gamma_i, \omega_i\right) = f\left(y_{ij1}, \ldots, y_{ijn_{ij}}\middle|\gamma_i, \omega_i\right)$ into product of univariate conditional densities, the resulting likelihood for the ith unit is

$$\begin{aligned} h\left(\boldsymbol{y}_i\middle|\gamma_i, \omega_i\right) &= \prod_{j=1}^{n_i} f\left(\boldsymbol{y}_{ij}\middle|\gamma_i, \omega_i\right) \\ &= \prod_{j=1}^{n_i}\prod_{k=2}^{n_{ij}} f\left(y_{ijk} \mid y_{ij1}, \ldots, y_{ij(k-1)}, \gamma_i, \omega_i\right) f\left(y_{ij1} \mid \gamma_i, \omega_i\right) \end{aligned} \tag{B2}$$

The overall conditional log-likelihoodover i, j, and k is

$$\log L = \sum_{i=1}^{n} \log h\left(\boldsymbol{y}_i\middle|\gamma_i, \omega_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} ll_{ijk}. \tag{B3}$$

By the property of multivariate normal distribution of $\boldsymbol{y}_{ij} \left| \gamma_i, \omega_i \right.$, each individual element y_{ijk} in response vector $\boldsymbol{y}_{ij}$ given $\gamma_i, \omega_i, y_{ij(k-1)}, \ldots, y_{ij1}$, also has a univariate normal distribution. Hence, the log-likelihood $ll_{ij1} = \log f\left(y_{ij1} \left| \gamma_i, \omega_i \right.\right)$ is an $N(\mu_{ij1}, \sigma^2_{ij1})$ random variable with $\mu_{ij1} = x^{\mathrm{T}}_{ij1}\beta + \gamma_i$ and $\sigma^2_{ij1} = \sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij1}}$; the log-likelihood $ll_{ij2} = \log f\left(y_{ij2} \left| y_{ij1}, \gamma_i, \omega_i \right.\right)$ is an $N(\mu_{ij2}, \sigma^2_{ij2})$ random variable with $\mu_{ij2} = x^{\mathrm{T}}_{ij2}\beta + \gamma_i + \sigma^2_{\upsilon_{ij}} / \left(\sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij1}}\right)\left(y_{ij1} - x^{\mathrm{T}}_{ij1}\beta - \gamma_i\right)$ and $\sigma^2_{ij2} = \sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij2}} - \sigma^4_{\upsilon_{ij}} / \left(\sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij1}}\right)$; and the log-likelihood $ll_{ijk*} = \log f\left(y_{ijk*} \left| y_{ij1}, \ldots, y_{ij(k*-1)}, \gamma_i, \omega_i \right.\right)$ is a $N(\mu_{ijk*}, \sigma^2_{ijk*})$ random variable when $k* = 3, \ldots, n_{ij}$. When $k^* \geqslant 3$, thelog-likelihood, mean, and variance in general are as follows:

$$ll_{ijk*} = -0.5 \log\left(2\pi\sigma^2_{ijk*}\right) - 0.5(y_{ijk} - \mu_{ijk*})^2 / \sigma^2_{ijk*}, \tag{B4}$$

$$\mu_{ijk*} = x^{\mathrm{T}}_{ijk*}\beta + \gamma_i + \frac{\sigma^2_{\upsilon_{ij}}}{k_{k*-1}} \sum\nolimits_{m=1}^{k*-1} \left(\prod\nolimits_{k \neq m}^{k*-1} \sigma^2_{\varepsilon_{ijk}}\right)(y_{ijm} - x^{\mathrm{T}}_{ijm}\beta - \gamma_i), \tag{B5}$$

$$\sigma^2_{ijk*} = \sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ijk*}} - \frac{\sigma^4_{\upsilon_{ij}}}{k_{k*-1}} \left(\sum\nolimits_{m=1}^{k*-1} \prod\nolimits_{k \neq m}^{k*-1} \sigma^2_{\varepsilon_{ijk}}\right), \tag{B6}$$

where $k_{k*-1} = \sigma^2_{\upsilon_{ij}} \sum_{m=1}^{k*-1} \prod_{k \neq m}^{k*-1} \sigma^2_{\varepsilon_{ijk}} + \prod_{k=1}^{k*-1} \sigma^2_{\varepsilon_{ijk}}$ and $\sigma^2_{\varepsilon_{ijk}} = \exp(\boldsymbol{\omega}^T_{ijk}\boldsymbol{\tau} + \omega_i)$.

To reduce computational time, the current mean and variance can be computed from the components that have been computed in the previous parts. The means and variances of the first and second log-likelihood (ll_{ij1} and ll_{ij2}) are computed the same way as mentioned before, namely,

$$\begin{aligned}
&\mu_{ij1} = x^{\mathrm{T}}_{ij1}\beta + \gamma_i \ \text{ and } \ \sigma^2_{ij1} = \sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij1}}, \\
&\mu_{ij2} = x^{\mathrm{T}}_{ij2}\beta + \gamma_i + \sigma^2_{\upsilon_{ij}} / \left(\sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij1}}\right)\left(y_{ij1} - x^{\mathrm{T}}_{ij1}\beta - \gamma_i\right) \ \text{ and} \\
&\sigma^2_{ij2} = \sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij2}} - \sigma^4_{\upsilon_{ij}} / \left(\sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij1}}\right).
\end{aligned}$$

The mean and variance of the third log-likelihood (ll_{ij3}) are given by

$$\begin{aligned}
&\mu_{ij3} = x^{\mathrm{T}}_{ij3}\beta + \gamma_i + \left(\sigma^2_{\upsilon_{ij}} / k_2\right)\left[t_{2,1}\left(y_{ij1} - x^{\mathrm{T}}_{ij1}\beta - \gamma_i\right) + t_{2,2}\left(y_{ij2} - x^{\mathrm{T}}_{ij2}\beta - \gamma_i\right)\right], \\
&\sigma^2_{ij3} = \sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij3}} - \sigma^4_{\upsilon_{ij}} t_2 / k_2,
\end{aligned}$$

where $k_2 = \sigma^2_{\upsilon_{ij}} t_2 + l_2; l_2 = l_1\sigma^2_{\varepsilon_{ij2}}; t_2 = t_{2,1} + t_{2,2}; t_{2,1} = t_{1,1}\sigma^2_{\varepsilon_{ij2}}; t_{2,2} = l_1$. The mean and variance of the fourth log-likelihood (ll_{ij4}) are as follows:

$$\begin{aligned}
\mu_{ij4} = x^{\mathrm{T}}_{ij4}\beta + \gamma_i + \left(\sigma^2_{\upsilon_{ij}} / k_3\right)&\left[t_{3,1}\left(y_{ij1} - x^{\mathrm{T}}_{ij1}\beta - \gamma_i\right) + t_{3,2}\left(y_{ij2} - x^{\mathrm{T}}_{ij2}\beta - \gamma_i\right)\right. \\
&\left. + t_{3,3}\left(y_{ij3} - x^{\mathrm{T}}_{ij3}\beta - \gamma_i\right)\right], \sigma^2_{ij4} = \sigma^2_{\upsilon_{ij}} + \sigma^2_{\varepsilon_{ij4}} - \sigma^4_{\upsilon_{ij}} t_3 / k_3,
\end{aligned}$$

where $k_3 = \sigma^2_{\upsilon_{ij}} t_3 + l_3; l_3 = l_2\sigma^2_{\varepsilon_{ij3}}; t_3 = t_{3,1} + t_{3,2} + t_{3,3}; t_{3,1} = t_{2,1}\sigma^2_{\varepsilon_{ij3}}; t_{3,2} = t_{2,2}\sigma^2_{\varepsilon_{ij3}}; t_{3,3} = l_2$.

In general when $k* \geqslant 3$, the mean and variance of the $k*$th likelihood (ll_{ijk*}) are

$$\mu_{ijk*} = x_{ijk*}^{\mathrm{T}}\beta + \gamma_i + \left(\sigma_{\upsilon_{ij}}^2 \,/\, k_{k*-1}\right) \sum_{m=1}^{k*-1} t_{k*-1,m} \left(y_{ijm} - x_{ijm}^{\mathrm{T}}\beta - \gamma_i\right), \tag{B7}$$

$$\sigma_{ijk*}^2 = \sigma_{\upsilon_{ij}}^2 + \sigma_{\varepsilon_{ijk*}}^2 - \sigma_{\upsilon_{ij}}^4 t_{k*-1} \,/\, k_{k*-1}, \tag{B8}$$

where $k_{k*-1} = \sigma_{\upsilon_{ij}}^2 t_{k*-1} + l_{k*-1}; l_{k*-1} = l_{k*-2}\sigma_{\varepsilon_{ij(k*-1)}}^2; t_{k*-1} = \sum_{m=1}^{k*-1} t_{k*-1,m}; t_{k*-1,m} = t_{k*-2,m}\sigma_{\varepsilon_{ij(k*-1)}}^2$ (for $m = 1$ to $k*-2$) and $t_{k*-1,k*-1} = l_{k*-2}$. Here, we define $t_{1,1}$ as 1 and l_1 to be $\sigma_{\varepsilon_{ij1}}^2$. With this recursive formula, the conditional log-likelihood is easily programmed in SAS PROC NLMIXED using the general likelihood option.

Appendix C: PROC NLMIXED SAS code

Below is a sample of syntax necessary to run the three-level mixed-effects LS model for at most four answered random prompts per subject per day described in this article. The first 13 observations for subject 1 to 3 from a hypothetical dataset **ONE**, used to illustrate the construction of the conditional likelihoods is given in Appendix A, Table A1. All programming was done using SAS version 9.1.

In the above code, uppercase letters are used for SAS specific syntax and lowercase letters are used for user-defined entities. In terms of the variables used in this syntax, `y_1`, `y_2`, `y_3`, and `y_4` denote the outcomes; `x1_1`, `x1_2`, `x1_3`, and `x1_4` denote prompt-varying covariates; `x2` denotes a day-level covariate; and `x3` denotes a subject-level covariate. `SubjectID` is a subject identifier. The number of answered random prompts per subject per day is represented by `freq`. The subject-level random location effect is named γ_i (or `gamma`) and random scale effect is named ω_i (or `w`). The model for the conditional means and variances for the first, second, third, and fourth conditional log-likelihood are respectively summarized by μ_{ij1} (or `eta1`), μ_{ij2} (or `eta2`), μ_{ij3} (or `eta3`), and μ_{ij4} (or `eta4`) for the means and σ_{ij1}^2 (or `var1`), σ_{ij2}^2(or `var2`), σ_{ij3}^2(or `var3`), and σ_{ij4}^2 (or `var4`) for the variances. The regression coefficients $\boldsymbol{\beta}$ (i.e., β_0, β_1, β_2, β_3) are named, respectively, as `beta0`, `beta1`, `beta2`, and `beta3` for the intercept, `x1_1 (x1_2/x1_3/x1_4)`, `x2`, and `x3`. The four conditional log-likelihoods (ll_{ij1}, ll_{ij2}, ll_{ij3}, and ll_{ij4}) are labeled with `z1`, `z2`, `z3`, and `z4`, respectively. The BS variance is given by $\sigma_{\gamma_i}^2$ (or `varg`), with λ_0 (or `lbd0`) indicating the reference BS variance (i.e., the BS variance when the subject-level covariate `x3` equals 0), and λ_1 (or `lbd1`) characterizing how this variance varies with `x3`. The within-subject and between-day (WS-BD) variance is given by $\sigma_{\upsilon_{ij}}^2$ (or `varu`), with α_0 (or `alpha0`) indicating the reference WS-BD variance (i.e., the WS-BD variance when the covariates `x2` and `x3` all equal 0), and α_1 (or `alpha1`) and α_2 (or `alpha2`) characterizing how this variance vary with day-level covariate `x2` and subject-level covariate `x3` correspondingly. For the model of the within-subjects within-day (WS-WD) variance for the first (second/third/fourth) response, `vare1(vare2/vare3/vare4)` is modeled in terms of a reference variance τ_0 (or `tau0`), with coefficients τ_1 (or `tau1`), τ_2 (or `tau2`), and τ_3 (or `tau3`) specified for the WS-WD variance that influences `x1_1(x1_2/x1_3/x1_4)`, `x2`, and `x3`, respectively. Here, `vare1` to `vare4` represent the WS-WD variances for the first to fourth responses, namely, $\sigma_{\varepsilon_{ij1}}^2$ to $\sigma_{\varepsilon_{ij4}}^2$.

Users must provide starting values for all parameters on the PARMS statement. To do so, it is beneficial to run the model in stages using estimates from a prior stage as starting values and setting the additional parameters to zero or some small value. For example, one can start by estimating a random-intercepts model with fixed effects ($\boldsymbol{\beta}$), BS variance λ_0 (or `lbd0`), WS-BD variance α_0 (or `alpha0`) and WS-WD variance τ_0 (or `tau0`) on log-scale. Estimates of these parameters can then be specified as starting values in a random scale model with the additional parameters σ_{ω}^2(or `varw`) and $\sigma_{\gamma\omega}$ (or `covgw`) being estimated. Finally, the full model with all other parameters λ_1 (or lbd1), α_1 (or alpha1), α_2 (or alpha2), τ_1 (or tau1), τ_2 (or tau2), and τ_3 (or tau3) can be estimated. In practice, this approach works well with `PROC NLMIXED`, which sometimes has difficulties in converging to a solution for complex models. Also, for complex models, it is sometimes the case that the default convergence criterion is not strict enough. In the above syntax, the convergence criteria is specified as `GCONV=1e-13` on the `PROC NLMIXED` statement.

Finally, with minor modification, the proposed SAS program can also fit a three-level LS model that has no random scale effect in the error variance, but still allows covariates to influence the variance of random intercept at each of the three levels, using a log-linear representation. This can be done

```
PROC NLMIXED DATA=ONE GCONV=1e-13;
PARMS beta0 = 6.9 beta1 = 0   beta2 =0 beta3=0
      lbd0  = 0.2 lbd1  = 0
      alpha0=-1.2 alpha1= 0   alpha2=0
      tau0  = 0.7 tau1  = 0   tau2  =0 tau3 =0
      covgw = 0   varw  = 0.6;
BOUNDS varw>0;

*****Common for all conditional logLik;
varg = EXP(lbd0 + lbd1*x3);
varu = EXP(alpha0 + alpha1*x2 + alpha2*x3);

*****1st conditional logLik;
xbeta1 = beta0 + beta1*x1_1 + beta2*x2 + beta3*x3;
vare1  = EXP(tau0 + tau1*x1_1 + tau2*x2 + tau3*x3 + w);
eta1 = xbeta1 + gamma;
var1 = varu + vare1;
z1 = -0.5*LOG(2*3.1415926*var1) - 0.5*(y_1-eta1)**2/var1;

*****2nd conditional logLik;
t1=1; l1=vare1; k1=varu*t1+l1;
p1 = y_1 - xbeta1 - gamma;
xbeta2 = beta0 + beta1*x1_2 + beta2*x2 + beta3*x3;
vare2  = EXP(tau0 + tau1*x1_2 + tau2*x2 + tau3*x3 + w);
eta2 = xbeta2 + gamma + (varu/k1)*(p1);
var2 = varu + vare2 - varu**2*t1/k1;
z2  = -0.5*LOG(2*3.1415926*var2) - 0.5*(y_2-eta2)**2/var2;

*****3rd conditional logLik;
t21=t1*vare2; t22=l1;
t2=t21+t22; l2=l1*vare2; k2=varu*t2+l2;
p2 = y_2 - xbeta2 - gamma;
xbeta3 = beta0 + beta1*x1_3 + beta2*x2 + beta3*x3;
vare3  = EXP(tau0 + tau1*x1_3 + tau2*x2  + tau3*x3 + w);
eta3 = xbeta3 + gamma + (varu/k2)*( t21*p1+t22*p2);
var3 = varu + vare3 - varu**2*t2/k2;
z3 = -0.5*LOG(2*3.1415926*var3) - 0.5*(y_3-eta3)**2/var3;

*****4th conditional loglik;
t31=t21*vare3; t32=t22*vare3; t33=l2;
t3=t31+t32+t33; l3=l2*vare3; k3=varu*t3+l3;
p3 = y_3 - xbeta3 - gamma;
xbeta4 = beta0 + beta1*x1_4 + beta2*x2 + beta3*x3;
vare4  = EXP(tau0 + tau1*x1_4 + tau2*x2  + tau3*x3 + w);
eta4 = xbeta4 + gamma + (varu/k3)*(t31*p1+t32*p2+t33*p3);
var4 = varu + vare4 - (varu**2*t3)/k3;
z4 = -0.5*LOG(2*3.1415926*var4) - 0.5*(y_4-eta4)**2/var4;

*****Overall loglik;
IF freq=1 THEN ll=z1;
ELSE IF freq=2 THEN ll=z1+z2;
ELSE IF freq=3 THEN ll=z1+z2+z3;
ELSE IF freq=4 THEN ll=z1+z2+z3+z4;

*****Model statement;
MODEL  y_1 ~GENERAL(ll);
RANDOM gamma w~NORMAL([0,0],[varg,covgw,varw]) SUBJECT=SubjectID;
RUN;
```

Statistics in Medicine

Table A1. First 13 observations of illustrated dataset, one that is simulated from 100 subjects.

	Subject	Day	Response				Covariates							Conditional Log-likelihood				Sum
Obs	**ID**	**ID**	**y_1**	**y_2**	**y_3**	**y_4**	**x1_1**	**x1_2**	**x1_3**	**x1_4**	**x2**	**x3**	**freq**					
1	1	1	9.7	8.3	10.0	—	1	0	1	—	0	1.3	3	ll_{111}	$+ll_{112}$	$+ll_{113}$		ll_{11}
2	1	2	11.0	—	—	—	0	—	—	—	1	1.3	1	ll_{121}				ll_{12}
3	1	3	6.9	7.5	—	—	1	0	—	—	0	1.3	2	ll_{131}	$+ll_{132}$			ll_{13}
4	1	4	11.1	4.7	8.5	—	1	1	0	—	0	1.3	3	ll_{141}	$+ll_{142}$	$+ll_{143}$		+) ll_{14}
																		ll_1
5	2	1	7.3	6.6	8.4	6.6	0	1	0	1	0	0.1	4	ll_{211}	$+ll_{212}$	$+ll_{213}$	$+ll_{214}$	ll_{21}
6	2	2	8.9	6.8	6.8	—	0	1	0	—	1	0.1	3	ll_{221}	$+ll_{222}$	$+ll_{223}$		ll_{22}
7	2	3	5.8	6.8	—	—	1	0	—	—	1	0.1	2	ll_{231}	$+ll_{232}$			ll_{23}
8	2	4	6.2	—	—	—	0	—	—	—	0	0.1	1	ll_{241}				+) ll_{24}
																		ll_2
9	3	1	8.0	—	—	—	0	—	—	—	1	0.0	4	ll_{311}				ll_{31}
10	3	2	7.9	—	—	—	0	—	—	—	0	0.0	3	ll_{321}				ll_{32}
11	3	3	8.0	8.0	5.6	—	1	1	0	—	0	0.0	2	ll_{331}	$+ll_{332}$	$+ll_{333}$		ll_{33}
12	3	4	7.0	7.2	8.8	8.3	1	0	0	0	1	0.0	1	ll_{341}	$+ll_{342}$	$+ll_{343}$	$+ll_{344}$	
							0	1	1	0	0	0.0						ll_{34}
13	3	5	7.1	8.2	7.0	8.3							4	ll_{351}	$+ll_{352}$	$+ll_{353}$	$+ll_{354}$	+) ll_{35}
																		ll_3

Conditional log-likelihood for the ith subject on jth day:

$$ll_{ij1} = \log f(y_{ij1} \mid \gamma_i, \omega_i) \qquad ll_{ij2} = \log f(y_{ij2} \mid y_{ij1}, \gamma_i, \omega_i) \qquad i = 1, \ldots, 100$$

$$ll_{ij3} = \log f(y_{ij3} \mid y_{ij2}, y_{ij1}, \gamma_i, \omega_i) \qquad ll_{ij4} = \log f(y_{ij4} \mid y_{ij3}, y_{ij2}, y_{ij1}, \gamma_i, \omega_i) \qquad j = 1, \ldots, n_i (1 \leqslant n_i \leqslant 5)$$

Overall conditional log-likelihood over 100 subjects, all days and all observations:

$$\log L = \sum_{i=1}^{100} \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} ll_{ijk} = ll_1 + ll_2 + \ldots + ll_{100}, \text{ where } ll_i = \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} ll_{ijk}.$$

easily by removing `covgw` **= 0** from `PARMS` statement, deleting the BOUNDS statement, dropping the variable `w` from `vare1-vare4`, and also modifying the random statement as: `RANDOM gamma w~NORMAL(`**0**`,varg)`

Acknowledgements

This work was supported by National Cancer Institute grants 5PO1 CA98262 and R21 CA140696. The authors would like to thank Professor Robin Mermelstein for use of the example data.

References

1. Stone A, Shiffman S. Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine* 1994; **16**:199–202.
2. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. *Journal of Happiness Studies* 2003; **4**:35–52.
3. De Vries M. *The experience of psychopathology: Investigating mental disorders in their natural settings*. Cambridge University Press: New York, 1992.
4. Bolger N, Davis A, Rafaeli E. Diary methods: capturing life as it is lived. *Annual Review of Psychology* 2003; **54**:579–616.
5. Walls TA, Schafer JL. *Models for intensive longitudinal data*. Oxford University Press: New York, 2006.
6. Schwartz JE, Stone A. The analysis of real-time momentary data: A practical guide. In *The science of real-time data capture: Self-report in health research*, Stone AA, Shiffman SS, Atienza A, Nebeling L (eds). Oxford University Press: Oxford, England, 2007.
7. Cleveland WS, Denby L, Liu C. Random scale effects. *Technical report, Bell Lab*, 2002. Available from: http://cm.bell-labs.com/cm/ms/departments/sia/wsc/webpapers.html.
8. Cleveland WS, Denby L, Liu C. Random location and scale effects: model building methods for a general class of models. *Computing Science and Statistics* 2002; **32**:3–10.
9. Hedeker D, Mermelstein RJ. Mixed-effects regression models with heterogeneous variance: Analyzing ecological momentary assessment data of smoking. In *Modeling Contextual Effects in Longitudinal Studies*, Little TD, Bovaird JA, Card NA (eds). Lawrence Erlbaum Associates: Mahwah, NJ, 2007; 183-206.
10. Hedeker D, Berbaum M, Mermelstein R. Location-scale models for multilevel ordinal data: Between- and within-subjects variance modeling. *Journal of Probability and Statistical Science* 2006; **4**:1–20.
11. Hedeker D, Demirtas H, Mermelstein RJ. A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Statistics and its Interface* 2009; **2**:391–401.
12. Hedeker D, Mermelstein RJ, Demirtas H. An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics* 2008; **64**:627–634.
13. Harvey AC. Estimating regression models with multiplicative heteroscedasticity. *Econometrica* 1976; **44**:461–465.
14. Aitkin M. Modeling variance heterogeneity in normal regression using GLIM. *Applied Statistics* 1987; **22**:332–339.
15. Fowler K, Whitlock MC. The distribution of phenotypic variance with inbreeding. *Evolution* 1999; **53**:1143–1156.
16. Leonard T. A Bayesian approach to the linear model with unequal variances. *Technometrics* 1975; **17**:95–102.
17. Ren `o R, Rizza R. Is volatility lognormal? Evidence from Italian futures. *Physica A: Statistical Mechanics and its Applications* 2003; **322**:620–628.
18. Shenk TM, White GC, Burnhamb KP. Sampling-variance effects on detecting density dependence from temporal trends in natural populations. *Ecological Monographs* 1998; **68**:445–463.
19. Vasseur H. Prediction of tropospheric scintillation on satellite links from radiosonde data. *IEEE Transactions on Antennas and Propagation* 1999; **47**:293–301.
20. Berndt E, Hall B, Hall R, Hausman J. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* 1974; **3**:653–665.
21. Gibbons RD, Hedeker D. Random effects probit and logistic regression models for three-level data. *Biometrics* 1997; **53**:1527–1537.
22. Raman R, Hedeker D. A mixed-effects regression model for three-level ordinal response data. *Statistics in Medicine* 2005; **24**:3331–3345.
23. Gueorguieva RV. Comments about joint modeling of cluster size and binary and continuous subunit-specific outcomes. *Biometrics* 2005; **61**:862–867.
24. Demirtas H. Simulation-driven inferences for multiply imputed longitudinal datasets. *Statistica Neerlandica* 2004; **58**:466–482.
25. Neuhaus JM, Kalbfeisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998; **54**:638–645.
26. Snijdersa T, Bosker R. *Multilevel Analysis*. Thousand Oaks: California, 1999; 65–66.
27. Gwaltney CJ, Bartolomei R, Colby SM, Kahler CW. Ecological momentary assessment of adolescent smoking cessation: A feasibility study. *Nicotine & Tobacco Research* 2008; **10**:1185–1190.
28. O'Connell KA, Gerkovich MM, Cook MR, Shiffman S, Hickcox M, Kakolewski KE. Coping in real time: Using ecological momentary assessment techniques to assess coping with the urge to smoke. *Research in Nursing & Health* 1998; **21**:487–497.
29. Longford N. *Random Coefficient Models*. Clarendon Press: Oxford, UK, 1993.

30. Hedeker D, Mermelstein RJ, Berbaum ML, Campbell RT. Modeling mood variation associated with smoking: an application of heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction* 2009; **104**:297–307.
31. Myles JP, Price GM, Hunter N, Day M, Duffy SW. A potentially useful distribution model for dietary intake data. *Public Health Nutrition* 2003; **6**(5):513–519.
32. Deboeck PR. Estimating dynamical systems: Derivative estimation hints from Sir Ronald A. Fisher. *Multivariate Behavioral Research* 2010; **45**:725–745.
33. Oertzen T, Boker SM. Time delay embedding increases estimation precision of models of intraindividual variability. *Psychometrika* 2010; **75**:158–175.
34. Norman GJ, Velicer WF, Fava JL, Prochaska JO. Dynamic typology clustering within the stages of change for smoking cessation. *Addictive Behaviors* 1998; **23**:139–153.
35. Jahng S, Wood PK, Trull TJ. Analysis of affective instability in EMA: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods* 2008; **13**:345–375.

