

RESEARCH ARTICLE

Extending the mixed-effects model to consider within-subject variance for Ecological Momentary Assessment data

Rachel Nordgren¹  | Donald Hedeker²  | Genevieve Dunton^{3,4} | Chih-Hsiang Yang⁴

¹Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, Illinois

²Department of Public Health Sciences, University of Chicago, Chicago, Illinois

³Department of Psychology, University of Southern California, Los Angeles, California

⁴Department of Preventive Medicine, University of Southern California, Los Angeles, California

Correspondence

Rachel Nordgren, Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, IL 60612.
Email: rnordg2@uic.edu

Funding information

National Heart, Lung, and Blood Institute, Grant/Award Number: R01 HL121330

Ecological Momentary Assessment data present some new modeling opportunities. Typically, there are sufficient data to explicitly model the within-subject (WS) variance, and in many applications, it is of interest to allow the WS variance to depend on covariates as well as random subject effects. We describe a model that allows multiple random effects per subject in the mean model (eg, random location intercept and slopes), as well as random scale in the error variance model. We present an example of the use of this model on a real dataset and a simulation study that shows the benefit of this model, relative to simpler approaches.

KEYWORDS

complex variation, heteroscedasticity, log-linear variance, multilevel, random effects, variance modeling

1 | INTRODUCTION

With smartphones and other such technology having become ubiquitous, studies designed to capture intense periods of data collection (eg, Ecological Momentary Assessment [EMA]^{1,2} and experience sampling^{3,4}) have become a standard way to examine variation within individual subjects. The datasets collected, with relatively large numbers of subjects and observations per subject, can be generically referred to as “intensive” longitudinal data⁵ and contain sufficient information to allow detailed modeling not just of the mean, but also the variance.

Mixed-effects regression models^{6,7} (MRMs) have become a standard method for analyzing multilevel (observations within clusters) data. Such models typically include multiple random subject effects to account for the influence of each subject upon their observations (ie, observations within a subject will be more similar than observations from different subjects). These models generally assume that all observations have an equal error variance (ie, amount of spread from predicted mean value, also called within-subject [WS] variance).

Heteroscedastic MRMs have also been developed to allow modeling of the error variance in terms of covariates such as gender, age, etc, typically through a log-linear structure.^{8,9} However, even after including covariates to account for variation across observations, some subjects will usually have greater variation, and others smaller variation. For example,

some subjects have fairly stable levels of positive mood, while others vary quite a bit. Similar patterns can be seen in a variety of measures, including negative mood, pain measures, and levels of physical activity. For this, we can allow a subject-specific level of variation by including a random subject effect when modeling the WS variance, in addition to the random effects used to model the mean response. The concept of a mixed-effects location-scale (MELS) model was formally introduced by Cleveland et al,¹⁰ which also summarizes much of the historical work regarding heterogeneity in scale. In their framework, the need for random scale effects was assessed after the location part of a unit regression model had been fitted, by examining the residuals for normality and homogeneity (within groups). Using a maximum marginal likelihood approach, joint estimation of the location and variance submodels has been developed in the context of a random subject intercept (location) model.¹¹⁻¹³

Allowing for multiple random location effects (eg, random intercepts and slopes) is a natural extension of the MELS model, and several authors have developed such models using Bayesian estimation approaches.¹⁴⁻¹⁸ Here, we provide a similar extension of allowing multiple random location effects to the MELS model, however using a maximum marginal likelihood estimation approach. Specifically, we use numerical integration techniques to integrate over the multiple random (location and scale) effects and derive the first and second derivatives for a Newton-Raphson solution. Adaptive Gauss-Hermite quadrature is most commonly used; however, Monte Carlo integration can be used as well.¹⁹ This then provides analysts with a frequentist approach for this class of models. As in the standard MELS model, we model the WS variance as log-linear, including both covariates and a random subject scale effect. The latter is further allowed to be correlated with the random location effects. Advantages of the likelihood approach include an ability to avoid the specification of priors, a more deterministic solution, simpler determination of convergence, and perhaps a more familiar model summary with standard errors and confidence intervals instead of credible intervals. Other differences between the Bayesian and likelihood approaches are described in the Discussion section.

This article is organized as follows. In Section 2, we define the model formally and give further rationale for why this model is needed, including a pictorial example. In Section 3, we discuss the maximum likelihood (ML) approach for estimating the model parameters. In Section 4, we present simulation results that both show that this method can recover coefficients correctly and also that a simpler model without random scale applied to data with heterogeneity will produce bias and poor coverage for the WS variance parameters. In Section 5, we describe a real-life dataset and present the corresponding results. In Section 6, we discuss advantages and limitations of the proposed model and we conclude this paper.

2 | MIXED-EFFECTS LOCATION-SCALE MODEL

Consider the following MRM^{6,7} for the outcome variable y , of subject i ($i = 1, 2, \dots, N$ subjects) at occasion j ($j = 1, 2, \dots, n_i$ observations):

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{v}_i + \epsilon_{ij}, \quad (1)$$

where \mathbf{x}_{ij} is the $p \times 1$ vector of regressors (typically, including a “1” for the intercept as the first element) and $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of regression coefficients. The regressors can either be at the subject level, vary across occasions, or be interactions of subject-level and occasion-level variables. Here, \mathbf{z}_{ij} is the $r \times 1$ design vector for the r random subject location effects \mathbf{v}_i . In the case of a random intercept and slope model, \mathbf{z}_{ij} is a vector containing both “1” for the random intercept and the value of an occasion-level variable at occasion j . In this case, \mathbf{v}_i contains the random subject intercept and slope for subject i , which measure the influence of individual i on both the mean of his/her repeated mood assessments and also the specific effect of the occasion-level variable.

The population distribution of \mathbf{v}_i is usually assumed to be a normal distribution with zero mean and $r \times r$ variance-covariance matrix $\boldsymbol{\Sigma}_v$. For a random intercept and slope model, this is given as $\boldsymbol{\Sigma}_v = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. The errors ϵ_{ij} are also assumed to be normally distributed in the population with zero mean and variance σ_ϵ^2 , and independent of the random effects.

This standard mixed model assumes homogeneous errors across all subjects and observations. For intensive longitudinal data, we can extend this and allow covariates in our modeling of the error variance. As previously mentioned, we utilize a log-linear model, namely, $\sigma_{\epsilon_{ij}}^2 = \exp(\mathbf{w}'_{ij}\boldsymbol{\tau})$. The variance is subscripted by i and j to indicate that it is specific to that subject and occasion as it depends on the values of the covariates \mathbf{w}_{ij} (and their coefficients). The covariates could be

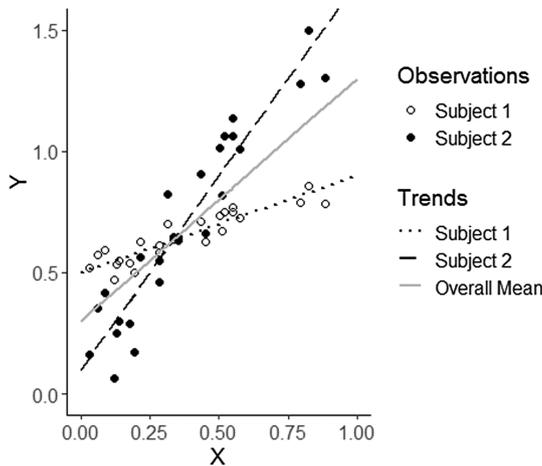


FIGURE 1 Plot of 2 example subjects and their trend lines

subject-level, occasion-level, or subject by occasion interactions, and these variables could be the same or different from those in the mean model. w_{ij} would usually include a (first) column of ones to include an intercept value, and if no further covariates are included, the model is equivalent to the earlier specification with $\exp(\tau_0) = \sigma_{\epsilon}^2$. The exponential function is used to ensure a positive value for any finite value of $w'_{ij}\tau$, and calculating the values of $\exp(\tau)$ yields variance ratios that allow easy interpretation for the effects of the covariates on the WS variance. Namely, for a particular variable w_k , $\exp(\tau_k)$ represents the ratio of error variances associated with a unit change of w_k .

We can further allow the WS variance to vary across subjects, above and beyond the contribution of covariates, namely,

$$\sigma_{\epsilon_{ij}}^2 = \exp(w'_{ij}\tau + \omega_i), \tag{2}$$

where the random subject (scale) effects ω_i are distributed in the population of subjects with mean 0 and variance σ_{ω}^2 . The idea for this is akin to the inclusion of the random (location) effects in Equation (1), namely, that the covariates do not account for all of the reasons that subjects differ from each other. The v_i subject-specific random effect values in (1) indicate how subjects differ in terms of their means and slopes and the ω_i values in (2) indicate how subjects differ in variation, beyond the effect of covariates. Taking the natural logarithm of both sides in (2) yields $\log(\sigma_{\epsilon_{ij}}^2) = w'_{ij}\tau + \omega_i$, which indicates that, if the distribution of ω_i is specified as normal, then the random effects are log normal subject-specific perturbations of the WS variance. Thus, the WS variances follow a log normal distribution at the individual level, and the skewed, nonnegative nature of the log normal distribution makes it a reasonable choice for representing variances.²⁰⁻²⁴ We can further generalize the model by allowing the location and scale random effects to be correlated.

Visually, Figure 1 presents the pertinent details of the model, in terms of a single continuous covariate. In this Figure, the average across all subjects is depicted with the solid gray line, and the average trends of two subjects are presented as a dotted line (Subject 1) and a dashed line (Subject 2). Hypothetical data points for these two subjects are also included in the plot. In a given dataset, there will be as many dashed lines as there are subjects, but, for simplicity, here, we only plot two subjects.

Relative to the overall line, the position of each dashed line when X is equal to zero is indicative of a person's random intercept location effect v_{1i} , which captures that subject's deviation from the mean response when X is zero. Also, relative to the overall line, the difference in slope of each dashed line is indicative of a person's random slope location effect v_{2i} , which captures how the slope of that subject deviates from the mean slope. In the plot, Subject 1 is (mostly) below and Subject 2 is (mostly) above the mean trend line. The heterogeneity in these lines is indicative of how much variance is observed across subjects: if the individual lines are close together then subjects are more similar, conversely if the individual lines are spread out then more heterogeneity is indicated.

The extent to which a person's data points vary around their line is indicative of a person's random scale effect ω_i . If the points are quite close to a subject's line, then that subject has low WS variance (eg, Subject 1). Conversely, if a subject's data points vary greatly around that person's line then more WS variation is indicated (eg, Subject 2).

Although a single covariate is shown in the plot, the general model given by (1) and (2) allows multiple covariates x and w to influence the location and scale through the coefficients β and τ .

3 | ESTIMATION

To simplify our calculations and improve computational stability, we reformulate our model to have standardized random effects (which are independent). For this, we can use the Cholesky factorization.²⁵ Below is the equivalence for two random effects.

$$\begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix} = \begin{bmatrix} T_1 & 0 \\ T_2 & T_3 \end{bmatrix} \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} = \begin{bmatrix} \sigma_{v_{1i}} & 0 \\ \sigma_{v_{1i}v_{2i}}/\sigma_{v_{1i}} & \sqrt{\sigma_{v_{2i}}^2 - \sigma_{v_{1i}v_{2i}}^2/\sigma_{v_{1i}}^2} \end{bmatrix} \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix}$$

Specifically, the random effects can be standardized using the Cholesky factorization of $\Sigma_v = TT'$. Here, $v_i = T\theta_i$, where θ_i follows a standard normal (bivariate) distribution. We can then write the mean model for the i th subject in vector form as

$$y_i = X_i\beta + Z_iT\theta_i + \epsilon_i. \tag{3}$$

Similarly, we can represent the scale random effect in standardized form, and as in the work of Hedeker and Nordgren,¹³ we can induce association between the location and scale random effects in the WS variance model

$$\sigma_{\epsilon_{ij}}^2 = \exp\left(\mathbf{w}'_{ij}\boldsymbol{\tau} + \theta_i\boldsymbol{\tau}^\theta + \sigma_\omega\vartheta_i\right). \tag{4}$$

Here, ϑ_i follows a standard normal distribution, and the coefficients $\boldsymbol{\tau}^\theta$ represent the association between a subject's location effects and the WS variance.

Based on our model assumptions, given values for the random effects, the residual for individual i , observation j , follows a Normal distribution with mean 0 and variance $\sigma_{\epsilon_{ij}}^2$. Furthermore, given the random effects, observations within a subject are independent. Referring back to Equations (3) and (4),

$$y_{ij} | \theta_i, \vartheta_i \sim N\left(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}T\theta_i, \exp\left(\mathbf{w}'_{ij}\boldsymbol{\tau} + \theta_i\boldsymbol{\tau}^\theta + \sigma_\omega\vartheta_i\right)\right). \tag{5}$$

We can obtain the joint likelihood by multiplying the conditional likelihood, symbolized by $f(y_i | \theta_i, \vartheta_i)$, with the likelihood of the random effects, symbolized by $g(\theta_i, \vartheta_i)$. After standardization, the random effects are independent and each follow a normal distribution with mean 0 and variance 1. Standard distribution theory allows us to get a marginal distribution by integrating one or more variables out of the joint distribution. Thus, in order to get the marginal likelihood for each subject (L_i), we integrate over all possible values for the random effects

$$L_i = \int_{\boldsymbol{\theta}_i, \vartheta_i} f(y_i | \boldsymbol{\theta}_i, \vartheta_i)g(\boldsymbol{\theta}_i, \vartheta_i)d\boldsymbol{\theta}_i d\vartheta_i. \tag{6}$$

The overall likelihood can then be obtained by summing the log of the likelihood of each subject. $\ell = \sum_{i=1}^N \ln L_i$. Following standard ML procedure, we can solve for the parameter values that maximize this log likelihood, using a numerical approximation for the integral. We have extended the MIXREGLS program¹³ to allow for slope random effects. The Supplementary Materials provides the derivation of the first and second derivatives for a Newton-Raphson solution. At convergence, the second derivative matrix can be inverted to obtain the large-sample variance-covariance matrix for the parameters. Additionally, one can use SAS PROC NLMIXED or the Stata package Merlin¹⁹ to produce these ML estimates and their standard errors. Examples of the syntax necessary for NLMIXED and Merlin are provided in the Appendix.

4 | SIMULATION RESULTS

4.1 | Examining parameter estimation

In order to compare the models estimated with and without random scale effects, we analyzed 500 successfully simulated datasets, each with 200 subjects with 25 observations each. Five hundred was chosen to allow an estimation error of 1% at 95% coverage, which was considered the primary outcome of interest.²⁶ The datasets were created following Equations (1) and (2) with the true values listed under Model 4 in Table 1. The true values as well as the summary statistics of x were chosen to be similar to an actual observed dataset. x contains one occasion-level variable (x_{WS}) and one subject-level (x_{BS})

TABLE 1 Simulation results

	True value	Mean(est)	Bias	std bias	sd(est)	RMSE	Width	Coverage
MODEL 1								
Fixed intercept	34.07	34.05	-0.02	-3.09	0.76	0.76	1.50	95%
Fixed x_{BS}	-8.56	-8.36	0.21	5.98	3.44	3.45	6.87	96%
Fixed x_{WS}	-1.76	-1.77	-0.01	-2.08	0.44	0.44	0.88	95%
Random intercept variance	45.38	44.93	-0.45	-9.18	4.93	4.95	9.53	93%
Random int/slope covariance	1.91	1.21	-0.70	-22.23	3.16	3.24	6.13	95%
Random slope variance	5.70	5.65	-0.05	-1.73	3.11	3.11	6.66	97%
WS variance intercept	4.36	4.52	0.16	365.68	0.04	0.17	0.04	0%
MODEL 2								
Mean intercept	34.07	34.05	-0.02	-2.59	0.76	0.76	1.50	95%
Mean x_{BS}	-8.56	-8.35	0.21	6.17	3.46	3.47	6.87	96%
Mean x_{WS}	-1.76	-1.77	-0.01	-2.40	0.45	0.45	0.87	95%
Random intercept variance	45.38	44.49	-0.89	-19.01	4.69	4.78	9.45	92%
Random int/slope covariance	1.91	1.57	-0.34	-10.86	3.15	3.16	6.12	95%
Random slope variance	5.70	6.23	0.53	14.48	3.68	3.72	6.34	86%
WS variance intercept	4.36	4.48	0.12	187.58	0.07	0.14	0.06	17%
WS variance x_{BS}	0.16	0.20	0.04	13.29	0.33	0.33	0.29	62%
WS variance x_{WS}	-0.10	-0.10	0.00	-0.63	0.07	0.07	0.12	89%
MODEL 3								
Mean intercept	34.07	34.06	-0.01	1.05	0.75	0.75	1.50	95%
Mean x_{BS}	-8.56	-8.46	0.10	2.93	3.48	3.48	6.92	95%
Mean x_{WS}	-1.76	-1.77	-0.01	-2.54	0.41	0.41	0.72	93%
Random intercept variance	45.38	44.93	-0.45	-9.17	4.92	4.94	9.52	93%
WS variance intercept	4.36	4.36	0.00	3.04	0.06	0.06	0.13	96%
WS variance x_{BS}	0.16	0.21	0.05	15.15	0.32	0.32	0.58	93%
WS variance x_{WS}	-0.10	-0.06	0.04	66.36	0.06	0.08	0.12	88%
Random scale Cholesky	0.52	0.51	-0.01	-38.97	0.03	0.04	0.07	92%
Random intercept/scale Cholesky	0.00	0.00	0.00	-4.22	0.04	0.04	0.09	96%
MODEL 4								
Mean intercept	34.07	34.05	-0.02	-2.79	0.76	0.76	1.50	95%
Mean x_{BS}	-8.56	-8.37	0.19	5.58	3.43	3.43	6.89	96%
Mean x_{WS}	-1.76	-1.77	-0.01	-2.24	0.44	0.44	0.85	95%
Random intercept variance	45.38	44.94	-0.44	-8.97	4.92	4.94	9.52	93%
Random int/slope covariance	1.91	1.70	-0.21	-6.97	2.95	2.96	5.71	94%
Random slope variance	5.70	5.64	-0.06	-2.02	3.19	3.19	5.32	85%
WS variance intercept	4.36	4.35	-0.01	-11.14	0.06	0.06	0.13	95%
WS variance x_{BS}	0.16	0.20	0.04	13.87	0.31	0.31	0.59	95%
WS variance x_{WS}	-0.10	-0.10	0.00	5.53	0.06	0.06	0.13	95%
Random scale Cholesky	0.52	0.50	-0.02	-40.63	0.05	0.05	0.10	94%
Random intercept/scale Cholesky	0.00	0.00	0.00	-2.19	0.04	0.04	0.09	95%
Random slope/scale Cholesky	0.00	0.00	0.00	2.80	0.11	0.11	0.43	97%

$x_{BS} = \bar{x}_i$ and $x_{WS} = x_{ij} - \bar{x}_i$.

Abbreviations: RMSE, Root Mean Squared Error; WS, within-subject.

variable. x_{BS} was simulated from a Beta distribution with parameters $\alpha = 1$ and $\beta = 5$, to represent a proportion with approximate mean 0.17 and standard deviation 0.14. A histogram is provided in the Supplementary Materials. Occasions were simulated using a Bernoulli distribution with $p = x_{BS}$, and x_{WS} was created by subtracting x_{BS} , so that the observations were mean-centered.

We estimated four separate mixed-effect models: (1) homogeneous errors, (2) WS variance covariates, (3) WS variance covariates plus a random scale effect, but without a random slope, and (4) all three random effects (ie, random intercept, slope on x_{WS} , and scale) and WS variance covariates. Note that Table 1 lists the results in terms of the Cholesky elements, although for interpretation purposes, they could be transformed to yield a variance-covariance matrix.

From the results, we calculated means and standard deviations (also referred to as the empirical standard errors) of the estimates of each parameter, along with whether the true value would be contained in a 95% Wald confidence interval, and the average width of that Wald confidence interval. Note that a Wald confidence interval is not truly appropriate

for variance parameters—a mixture of chi-squared distributions should be used—leading to smaller expected coverages, especially for dataset with smaller numbers of subjects.²⁷ Bias is defined as the difference between the mean estimate and the true value, but that does not consider the spread of the estimates. There are two ways that this information is combined. Standardized Bias is bias divided by the standard deviation of the estimates, times 100. A value of less than 50 is usually considered not of concern.²⁸ For this index, bias is considered less concerning if the spread of the estimates is larger. Root Mean Squared Error (RMSE) is the square root of the bias squared plus the variance of the estimates. For this index, smaller spread of the estimates leads to smaller RMSE, as does smaller bias.

Examining the results for the first model, results are reasonable for all parameters except the last, which is the log of the error variance. This value has a large Standardized Bias (366) and a nonexistent coverage (0). This is not surprising since this model assumes homogeneous errors, and the simulated data had heterogeneous errors.

From the results for the second model, results again are reasonable for the first six parameters (three fixed effects and three variance/covariances of random location effects). The log of error variance when the covariates are zero still has a large Standardized Bias (188) and a low coverage (17%). One could make the case that the true value for this parameter for Models 1 and 2 also includes half the variance of the (omitted) scale random effect (since the random scale effect follows a log-normal distribution) and thus would be $4.36 + 0.52^2/2 = 4.50$. Using this value would decrease the absolute bias to 0.02 for both Models 1 and 2, indicating that these models were estimated well, they are just the incorrect model (we chose to list the same true value across all models, since it is equal to the parameter value used when the datasets were created). In addition, the coverage for the effect of x_{BS} on scale is too small (62%), indicating that the standard error estimated is too small. This agrees with results reported by Leckie et al¹⁵ who found similar coverage problems for WS variance covariates, especially subject-level ones, if random scale was present but omitted from the model.

Model 3 does not include a random slope effect and so does not also include covariances between the random slope effect and the other random effects. Unsurprisingly, the effects are seen in the coefficients for x_{WS} , since that is the parameter that should have a random slope effect. The width of the effect of x_{WS} on location is artificially small relative to all other models (0.72 vs 0.87). The bias of the effect of x_{WS} on WS variance is much larger than the other models, and the Standardized Bias (66) is concerning, suggesting that the lack of a random slope leads to an inflated estimated effect on WS variance for that variable. The width for the standard deviation of the random scale effect is small, leading to a smaller coverage (92%).

The final model includes all three random effects and has fixed the coverage and bias issues of the Models 1 and 2. If we examine the widths for the scale intercept and BS covariate, the values for Model 4 are approximately double those of Model 2 (0.13 vs 0.06 and 0.59 vs 0.29). Further, the width for the standard deviation of the random scale effect has increased from that of Model 3 (0.10 vs 0.07), and coverage increases to 94%. The bias of the WS variance intercept is now close to zero, with the extra variation being correctly modeled with the random effect at the individual level. We may note that the coverage for the random slope variances is lower than 95%; this is a likely a reflection of the fact that the use of Wald confidence intervals for variances is not ideal.

The results in Table 1 are for 500 simulated datasets in which convergence was achieved for all four models. Approximately 14% of simulated datasets did not properly converge for all models, and the most common reason was inability to estimate a nonzero slope variance, accounting for 72% of the non-convergent datasets (10% of all simulated datasets). Further, examining those simulation results that did converge but did not cover the true value for the random slope variance, nearly all had an estimate closer to 0, with 61% of those less than 0.5 and 95% less than 1.5 (true value 5.7). This difficulty was likely caused in large part by the decision to use a binary variable for the random slope, as well as a relatively small mean for the proportion of 1s (0.17, which was based on the real dataset presented later).

4.2 | Examining the amount of data required for accurate estimation

To get a sense of how many subjects and observations are required to estimate the parameters of this model, we performed two additional simulation studies, one with 100 subjects and one with 300 subjects. We experimented with a variety of observations per subject, aiming for a convergence rate of at least 90%, and chose 75 observations and 25 observations, respectively. Results for Model 4 across 200 repetitions, along with 200 repetitions from the previous simulation study, are shown in Supplementary Tables 1 and 2. Two hundred repetitions were a compromise necessary because of the complicated nature of the estimation, but the estimation error at 95% coverage only increases to 1.5% with this number of repetitions.²⁶

Overall, with 300 subjects and 25 observations, convergence was achieved 90% of the time. Relative to the original setup, it was more consistent (smaller $est(sd)$) in estimating all parameters and had a smaller width for all parameters except the Cholesky for random slope and scale covariance. In some cases, this smaller width caused the coverage to decrease; this

would presumably be improved if more replications were performed. A similar proportion of simulated datasets allowed convergence of Model 1 but failed for Model 4. The proportion of datasets failing to estimate a nonzero slope reduced to 5% (as compared to 10% for 200 subjects with 25 observations).

With 100 subjects and 75 observations, the convergence rate was 94%. Relative to 200 subjects with 25 observations, it was more consistent (smaller $\text{est}(\text{sd})$ and width) in estimating quantities within each subject, most notably the random slope variance and the effect of x_{WS} on both location and scale. Estimation of the random scale standard deviation remained consistent but had better coverage, and only 2% of simulated datasets converged for Model 1 but not for Model 4 (as compared to $14 - 10 = 4\%$ for the original setup). Not only did the coverage for the random slope variance increase, but the proportion of datasets failing to estimate a nonzero slope reduced to 4% (as compared to 10%), and datasets that did not cover the true random slope variance had more varied estimates, with only 31% of those less than 0.5 and 38% less than 1.5 (true value 5.7). With only 100 subjects, there was less accuracy in estimating certain population level quantities, most notably the random location intercept variance, the mean intercept, and the effect of x_{BS} on location. This suggests that if there is interest in accurate estimation of these population level quantities, even with 75 observations/subject, more than 100 subjects are probably necessary.

In summary, the method of estimation proposed in this paper does a good job of recovering the true parameter values. Furthermore, we can observe that the first two models overestimate the WS scale intercept parameter and underestimate the standard error for WS scale parameters, while Model 3 underestimates the standard error for the WS fixed parameter, both of which could lead to false conclusions concerning the effects of the WS variable.

5 | EXAMPLE: MOOD AND SOCIAL CONTEXT DATASET

We consider a dataset from Project MATCH,²⁹ where EMA data collection was performed on mother-child dyads. In the positive affect section, questions were asked relating to happiness, joyfulness, and calmness, rated from 1 to 4, and the available responses were averaged and multiplied by 10. At each prompt, subjects also answered whether they were alone (1 = alone, 0 = not alone).

We consider the data for the mothers at baseline, with 195 subjects and 4386 observations, so that the mean number of observations per subject is 22.49 (range 1-34). The positive mood outcome variable had a mean of 26.51 and a standard deviation of 7.62. Maternal age was centered at the mean (41.13 years, $\text{sd} = 6.1$ years) and divided by 10, so that one unit corresponds to 10 years removed from the mean. We decomposed the variable alone in two parts—the proportion of observations that subject was alone, and the difference between the given observation and the subject mean,³⁰ namely, $\tilde{x}_{ij} = \bar{x}_i + (x_{ij} - \bar{x}_i)$. This was done to allow separate modeling of the between-subject effect (the effect of the tendency for a subject to be alone), and the WS effect (the effect of being alone at a given occasion). The subject means for alone ranged from 0.00 to 0.82, with a mean of 0.18 and a standard deviation of 0.15. We also considered whether each observation occurred on the weekend (of which 0.48 of them did), and allowed the effect of weekend, relative to weekday, to vary across subjects by including a random slope.

Table 2 presents the results for several models with increasing complexity. To compare models, we can use likelihood-ratio tests to determine if the simpler models are significantly worse than the more complicated ones, with the caveat that testing for variance components requires a mixture of two chi-squared distributions,²⁷ so two degree of freedom values are specified. Between Model 0 and Model 1 (random slope effect added), $\chi^2 = 26.49$, $df = 1, 2$, and $p < 0.01$, suggesting that Model 0 is significantly worse and indicating that there is significant heterogeneity in the slope for weekend, relative to weekday, on positive affect. Between Model 1 and Model 2 (fixed scale effects added), $\chi^2 = 8.66$, $df = 4$, and $p = 0.07$, indicating a marginal improvement by adding in the covariates for the WS variance. Comparing Model 2 and Model 4 (random scale effect added), $\chi^2 = 216.22$, $df = 2, 3$, and $p < 0.01$, which rejects Model 2 and indicates that there is a great deal of heterogeneity across subjects in terms of the random scale. Comparing Models 3 and Model 4 (random slope effect added), $\chi^2 = 20.71$, $df = 2, 3$, and $p < 0.01$, rejects Model 3 and affirms the need for the random slope, over and above the inclusion of the random scale.

Examining the significant mean effects in Model 4, we can draw the following conclusions. Prompts on a weekend had a higher mean positive mood by 1.57 points. Subjects who had a higher fraction of prompts when alone had on average lower positive mood of 5.44 points. Within each subject, prompts where the subject was alone had a lower mean positive mood by 1.01 points. Age did not have a significant effect on the mean.

Note that, in these models, the Cholesky terms have been converted into variances and covariances. The specific variable transformation is shown in the Appendix. There was quite a bit of subject heterogeneity for positive mood, with a 95%

TABLE 2 Results for example dataset across five nested models: maximum likelihood estimates (standard errors), with Cholesky terms converted to variance/covariance elements

Variable	Model 0	Model 1	Model 2	Model 3	Model 4
MEAN COEFFICIENTS					
Intercept	26.78 (0.52)	26.77 (0.52)	26.76 (0.52)	26.88 (0.52)	26.79 (0.53)
Weekend	1.55 (0.19)	1.53 (0.24)	1.54 (0.24)	1.40 (0.17)	1.57 (0.23)
Age	-0.51 (0.53)	-0.50 (0.53)	-0.49 (0.53)	-0.50 (0.54)	-0.49 (0.53)
Alone BS	-5.38 (2.14)	-5.26 (2.14)	-5.26 (2.14)	-5.48 (2.17)	-5.44 (2.17)
Alone WS	-1.11 (0.26)	-1.08 (0.26)	-1.06 (0.25)	-1.03 (0.23)	-1.01 (0.23)
RANDOM LOCATION EFFECTS					
Intercept variance	18.61 (2.10)	19.79 (2.37)	19.68 (2.37)	19.12 (2.12)	19.96 (2.34)
Covariance		-2.19 (1.23)	-2.03 (1.22)		-1.51 (1.07)
Slope (weekend) variance		4.39 (1.13)	4.26 (1.17)		2.91 (0.99)
SCALE COEFFICIENTS (log-linear)					
Intercept	3.64 (0.02)	3.61 (0.02)	3.65 (0.04)	3.53 (0.07)	3.51 (0.08)
Weekend			-0.07 (0.05)	-0.06 (0.05)	-0.06 (0.05)
Age			-0.06 (0.04)	-0.11 (0.07)	-0.11 (0.07)
Alone BS			-0.07 (0.15)	0.14 (0.30)	0.15 (0.30)
Alone WS			-0.12 (0.06)	-0.13 (0.06)	-0.14 (0.07)
RANDOM SCALE EFFECT					
Intercept variance				0.27 (0.04)	0.28 (0.04)
Covariance with location intercept				-0.22 (0.21)	-0.34 (0.22)
Covariance with slope					0.17 (0.12)
DEVIANCE	28876.96	28850.47	28841.81	28646.30	28625.59

$$x_{BS} = \bar{X}_i \text{ and } x_{WS} = x_{ij} - \bar{X}_i$$

Abbreviations: BS, between-subject; WS, within-subject.

plausible values interval³¹ for the intercept in Model 4 of $26.79 \pm 1.96 \times \sqrt{19.96} = (18.03, 35.55)$. Significant heterogeneity for the effect of weekend can be seen both in the significant random slope variance as well as the superiority of Model 1 vs Model 0 and of Model 4 vs Model 3, and a 95% plausible values interval in Model 4 for the slope of weekend of $1.57 \pm 1.96 \times \sqrt{2.91} = (-1.77, 4.91)$. The covariance between the intercept and slope was somewhat negative, suggesting that those subjects with larger random intercept values had smaller or more negative random slope values, but it was not statistically significant. Further, there was a large amount of subject heterogeneity for scale of positive mood, and the random scale effect was effectively independent of the location random effects, as the random location effects had no significant effects on the WS variance.

In terms of the covariate effects on the WS variance (scale effects), within the same subject, subjects had a smaller amount of variation in mood when alone, by $\exp(-.14) = 0.87$. A less obvious change from Model 2 to Models 3 and 4 is that the standard errors for the scale coefficients of the subject-level covariates (age and fraction alone) approximately double with the addition of the random scale effect. Thus, in agreement with the simulation study and the aforementioned results in the work of Leckie et al,¹⁵ inference about scale covariates can be misleading unless the random scale is included in the model. Model 4 also has a much smaller slope variance than Model 2, suggesting that the variation modeled as a random scale effect in Model 4 was partially absorbed by the random slope variance in Model 2. Given that both variances are at the subject level, this suggests that ignoring scale variance, if present in the data, can lead to an over-estimation of slope heterogeneity.

To examine individual subjects, one may be interested in estimated values of the random intercept, slope, and scale effects. A standard choice for this is to use the expected “a priori” or empirical Bayes estimator.³² In this example, of interest are the subjects most affected by the weekend variable, which are those who have a large value for the random slope effect. Further, the subject’s empirical Bayes estimate of the random scale gives a measure of how well the observations are fit by the model. In Figure 2, we provide dot plots of the observed positive affect responses for four subjects from the dataset. In Figure 2A, the subject has a large and positive random slope estimate and a small random scale estimate. This subject has a much higher positive affect mean on weekend days, relative to weekdays, and both sets of observations are relatively tightly clustered. In Figure 2B, the subject has a large and negative random slope estimate and a small random scale estimate. This subject also provides consistent responses, but has a higher mean response on the weekdays, relative to the weekend. In Figure 2C, the subject has a large and positive random slope estimate and a large random scale estimate. As in

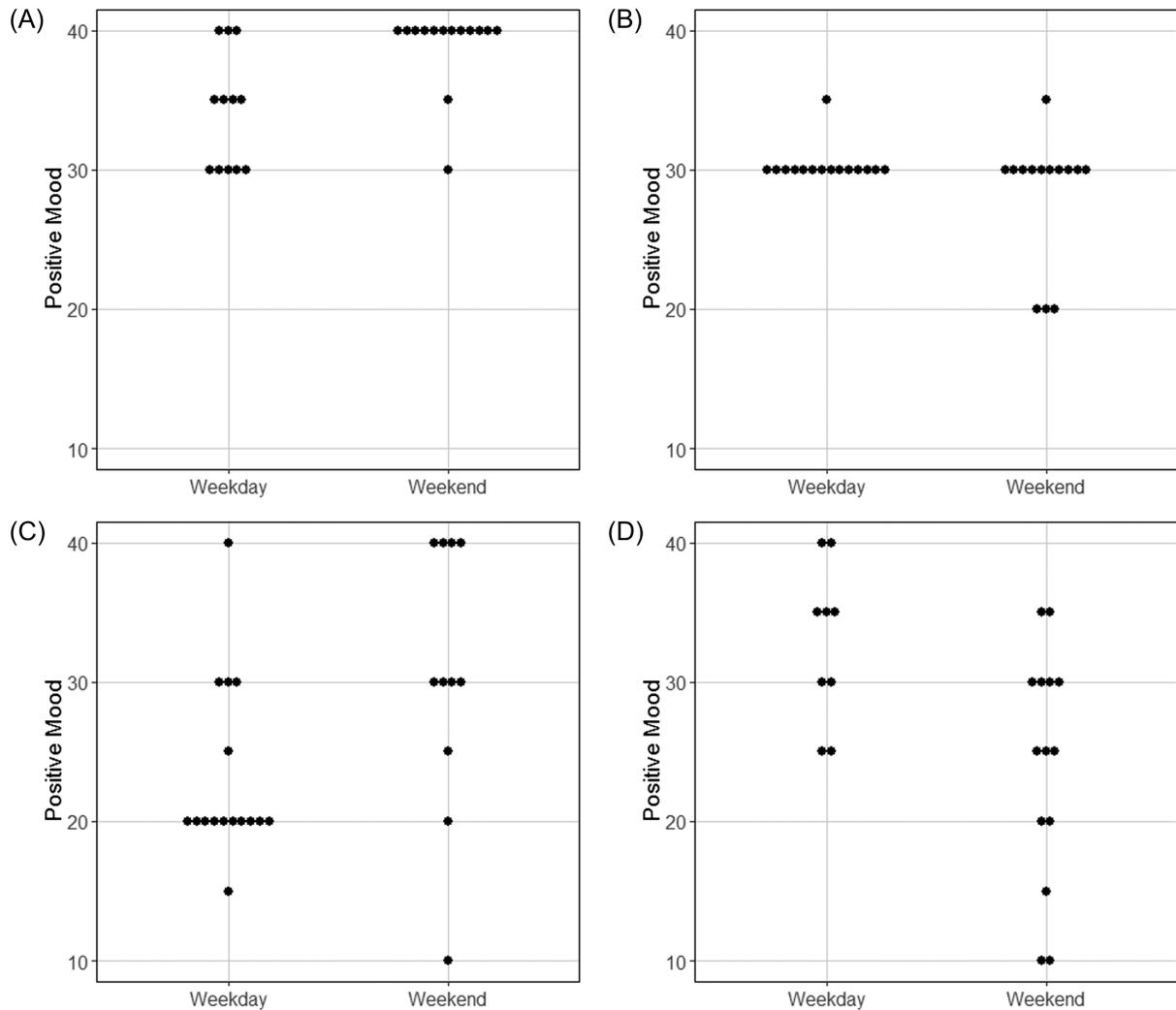


FIGURE 2 Dot plots by weekend/weekday for 4 example subjects. x-axis is weekday/weekend, and y-axis is positive mood. In A) observations have a greater mean on the weekend days and are relatively tightly clustered. In B) observations have a lower mean on the weekend days and are relatively tightly clustered. In C) observations have a greater mean on the weekend days but are relatively diffuse. In D) observations have a lower mean on the weekend days but are relatively diffuse

Figure 2A, the weekend observations have a higher mean value, but for this subject, the observations are quite diffuse. In Figure 2D, the subject has a large and negative random slope estimate and a large random scale estimate. As in Figure 2B, the weekday observations have a higher mean value, and like Figure 2C, the responses are erratic. This ability to screen for subjects of particular interest can allow further exploration of what makes that subject unusual. For example, it might be interesting to see if subjects with a negative random slope worked weekends or had some other reasons to have lower positive affect on the weekends.

6 | DISCUSSION

In this paper, we have described a model for intensive longitudinal data that includes multiple random location effects, in addition to a random scale effect. Such data often arise from EMA studies, in which the ability of model both the mean and variance structure of the data is of interest. Relative to a more standard mixed model, our approach allows covariates to influence the WS variance and includes a random subject scale effect. Thus, subjects are allowed to vary in their consistency/erraticism over and above the influence of covariates. Furthermore, the random location effects (eg, intercept and slope) and the random scale effect can be correlated. We have conducted a small simulation study that verifies that the model parameters can be reliably estimated and have provided an analysis of an existing EMA dataset. The simulation study indicated that models with random scale may be significantly better than those without and that

relying upon simpler models without a random scale effect may produce biased standard errors for scale parameters if a subject's scale varies beyond that which can be explained by covariates. This is important because some software programs (eg, SAS PROC MIXED, Stata command mixed, MLwiN, and HLM) do allow covariates to influence the WS variance, but do not include a random scale effect. Our simulation suggests that, without the random scale effect, the inference of such covariate effects may be misleading. A similar result for clustered data was observed in the work of Leckie et al.¹⁵

The approach described in this paper uses a full likelihood approach, in contrast to previous papers that have used a Bayesian approach for similar and/or related models.¹⁴⁻¹⁸ Given the flexibility of Bayesian software, all of these papers allow for multiple random location and scale effects, and Goldstein et al¹⁷ additionally allow for a nonlinear growth model across time. While focused on a linear model, we provide a similar extension of allowing multiple random location effects to the MELS model, however using a ML estimation approach. Specifically, we use numerical integration techniques to integrate over the multiple random (location and scale) effects and derive the first and second derivatives for a Newton-Raphson solution. This then provides analysts with a frequentist approach for this class of models. As in the standard MELS model, we model the WS variance as log-linear, including both covariates and a random subject scale effect. The latter is further allowed to be correlated with the random location effects.

This likelihood-based approach can be implemented using SAS PROC NLMIXED and the Stata program Merlin,¹⁹ which are described in the Appendix. Also, the first and second derivatives for a likelihood-based solution using the Newton-Raphson algorithm are included in the Supplementary Materials. Using these results, we have extended the MIXREGLS program¹³ to provide likelihood-based estimation for the model described in this paper, including multiple random location effects. In the Appendix, we provide information about accessing our program MIXREGLS, which was used to produce the results in this paper. In terms of differences between these likelihood-based procedures, all three programs use adaptive Gauss-Hermite quadrature, but Merlin also has the option to use Monte Carlo integration. Since MIXREGLS is optimized for the model presented in this paper, it is usually faster to converge than NLMIXED, which provides more general programming facilities for a wide range of mixed models. Additionally, users must provide starting values in NLMIXED, which are not required by MIXREGLS. A disadvantage of the quadrature approach is that the total number of quadrature points required increases exponentially with each additional random effect. Merlin, when using Monte Carlo integration option, avoids this exponential increase, although additional random effects can still increase the complexity of the estimation in some cases. Thus, Merlin can provide faster estimation, especially for models with multiple random effects. Additionally, both NLMIXED and Merlin are housed within general software packages (SAS and Stata, respectively) and so can more readily provide data management, plots, and post-estimation facilities.

Although there are advantages to the likelihood approach relative to the Bayesian approach (including an ability to avoid the specification of priors, a more deterministic solution, simpler determination of convergence, and standard errors rather than credible intervals), there are also disadvantages. In particular, it can be the case that computational issues arise in the iterative ML solution, which prevents the program from converging successfully. This can sometimes be the result of a poor specification of starting values, which is particularly germane for use of SAS PROC NLMIXED. In this regard, estimating the model parameters in stages can be helpful (for example, estimating the model with location random effects only, then adding fixed scale covariates, then adding the random scale effect) both for obtaining good starting values for the next stage and also for determining at which point the estimation may have computational difficulties. In our programming of MIXREGLS, we have found that overall model convergence is improved by this sequential fit of models of increasing complexity. In some situations, increasing the number of quadrature points can also help in estimation, though at the expense of increased computational time. In our use of MIXREGLS, we typically use 10 or 11 point adaptive quadrature, but sometimes need to increase this number. As observed in the simulation studies, increasing the number of subjects or the number of observations per subject or both improves the rate of convergence. Of course, this is typically not something that an analyst has control over with a real dataset. Ultimately, in some cases, it may be that the data do not support estimation of the full model, and so, some simplification is necessary.

Our analysis of the EMA dataset provides some interesting information about the effect of being alone on positive mood. We found a strongly significant negative effect on mean positive mood for both prompts when the subject was alone and the overall fraction of prompts when alone. This matches previous effects reported in adolescents.^{13,33,34} We also found a significant decrease in positive mood variability when the subject was alone, but no significant effect on variability of the fraction of prompts when alone. This does not match the previous effects reported in other works^{13,33,34}—in adolescents, a higher fraction of prompts when alone led to a strongly significant increase in positive mood variability; while within each subject, being alone led to a marginal increase in variability.

The overall objective of the MATCH study was to examine the effect of maternal stress on child outcomes. To be eligible, women had to have an 8-12-year old child. It seems reasonable to assume that a sizable proportion of occasions when the

mother was not alone, she was either at work or dealing with her child(ren). Under either circumstance, it is not surprising that there might be greater variability in mood. In contrast, adolescents with a strong social network would have a lower fraction of prompts when alone as well as decreased variability in positive mood. Further, it is easy to imagine adolescents experiencing mood swings which are damped by the presence of others and more severe when alone.

Limitations of this model include a limit of one random scale effect and the assumption that the outcome variable is continuous. As mentioned above, several authors have proposed models that have included multiple random scale effects using Bayesian software.¹⁴⁻¹⁸ In terms of ordinal outcomes, we are investigating extending this model to handle such outcomes, by extending some of our previous work.^{33,35} This would allow greater flexibility in analyzing ordinal data, which are commonly found in many research areas, but requires more complicated estimation because of the need for a (non-linear) link function.

ACKNOWLEDGEMENT

The project described was supported by National Heart, Lung, and Blood Institute (grant number R01 HL121330). The content is solely the responsibility of the authors and does not necessarily represent the official views of National Heart, Lung, and Blood Institute.

ORCID

Rachel Nordgren  <https://orcid.org/0000-0003-3664-1208>

Donald Hedeker  <https://orcid.org/0000-0001-8134-6094>

REFERENCES

1. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. *J Happiness Stud.* 2003;4:35-52.
2. Stone A, Shiffman S. Ecological Momentary Assessment (EMA) in behavioral medicine. *Ann Behav Med.* 1994;16:199-202.
3. Scollon CN, Prieto C-K, Diener E. Experience sampling: promises and pitfalls, strengths and weaknesses. *J Happiness Stud.* 2003;4:5-34.
4. Barrett FL, Barrett DJ. An introduction to computerized experience sampling in psychology. *Soc Sci Comput Rev.* 2001;19:175-185.
5. Walls TA, Schafer JL. *Models for Intensive Longitudinal Data.* New York, NY: Oxford University Press; 2006.
6. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* New York, NY: Springer-Verlag; 2000.
7. Hedeker D, Gibbons RD. *Longitudinal Data Analysis.* New York, NY: Wiley; 2006.
8. Hedeker D, Berbaum M, Mermelstein R. Location-scale models for multilevel ordinal data: between- and within-subjects variance modeling. *J Probab Stat Sci.* 2006;4:1-20.
9. Hedeker D, Mermelstein RJ. Mixed-effects regression models with heterogeneous variance: analyzing ecological momentary assessment data of smoking. In: Little TD, Bovaird JA, Card NA, eds. *Modeling Ecological and Contextual Effects in Longitudinal Studies of Human Development.* Mahwah, NJ: Erlbaum; 2007.
10. Cleveland WS, Denby L, Liu C. Random location and scale effects: model building methods for a general class of models. In: Wegman EJ, Martinez YM, eds. *Computer Science and Statistics.* Vol 32. Fairfax Station, VA: Interface Foundations of North America; 2000:3-10.
11. 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.
12. Hedeker D, Mermelstein RJ, Demirtas H. Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statist Med.* 2012;31(27):3328-3336.
13. Hedeker D, Nordgren R. MIXREGLS: a program for mixed-effects location scale analysis. *J Stat Softw.* 2013;52(12):1-38.
14. Rast P, Hofer SM, Sparks C. Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using bugs/jags. *Multivar Behav Res.* 2012;47(2):177-200.
15. Leckie G, French R, Charlton C, Browne W. Modeling heterogeneous variance-covariance components in two-level models. *J Educ Behav Stat.* 2014;39(5):307-332.
16. Kapur K, Li X, Blood EA, Hedeker D. Bayesian mixed-effects location and scale models for multivariate longitudinal outcomes: an application to ecological momentary assessment data. *Statist Med.* 2015;34(4):630-651.
17. Goldstein H, Leckie G, Charlton C, Tilling K, Browne WJ. Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. *Stat Methods Med Res.* 2018;27(11).
18. Lin X, Hedeker D, Mermelstein RJ. A three level Bayesian mixed effects location scale model with an application to ecological momentary assessment data. *Statist Med.* 2018;37:2108-2119.
19. Crowther MJ. merlin - a unified modelling framework for data analysis and methods development in stata. arXiv:1806.01615. 2018.
20. Fowler K, Whitlock MC. The distribution of phenotypic variance with inbreeding. *Evolution.* 1999;53:1143-1156.
21. Leonard T. A Bayesian approach to the linear model with unequal variances. *Technometrics.* 1975;17:95-102.

22. Renò R, Rizza R. Is volatility lognormal? Evidence from Italian futures. *Phys A Stat Mech Appl*. 2003;322:620-628.
23. Shenk GC, White TM, Burnhamb KP. Sampling-variance effects on detecting density dependence from temporal trends in natural populations. *Ecological Monographs*. 1998;68:445-463.
24. Vasseur H. Prediction of tropospheric scintillation on satellite links from radiosonde data. *IEEE Trans Antennas Propag*. 1999;47:293-301.
25. Bock RD. *Multivariate Statistical Methods in Behavioral Research*. New York, NY: McGraw-Hill; 1975.
26. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. arXiv:1712.03198. 2017.
27. Verbeke G, Molenberghs G. The use of score tests for inference on variance components. *Biometrics*. 2003;59(2):254-62.
28. Demirtas H. Simulation-driven inferences for multiply imputed longitudinal datasets. *Statistica Neerlandica*. 2004;58:466-482.
29. Dunton GF, Liao Y, Dzibur E, et al. Investigating within-day and longitudinal effects of maternal stress on children's physical activity, dietary intake, and body composition: protocol for the match study. *Contemp Clin Trials*. 2015;43:142-154.
30. Neuhaus JM, Kalbfleisch JD. Between- and within- cluster covariate effects in the analysis of clustered data. *Biometrics*. 1998;54:638-645.
31. Raudenbush SW, Bryk AS. *Hierarchical Linear Models*. 2nd ed. Thousand Oaks, CA: Sage; 2002.
32. Bock RD, Aitken M. Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. *Psychometrika*. 1981;46:443-459.
33. Hedeker D, Demirtas H, Mermelstein RJ. A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Stat Interface*. 2009;2:391-402.
34. Li X, Hedeker D. A three-level mixed-effects location scale model with an application to ecological momentary assessment data. *Statist Med*. 2012;31(26):3192-3210.
35. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics*. 1994;50:933-944.
36. Leckie G. runmixregls: a program to run the MIXREGLS mixed-effects location scale software from within stata. *J Stat Softw*. 2014;59(2).

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Nordgren R, Hedeker D, Dunton G, Yang C-H. Extending the mixed-effects model to consider within-subject variance for ecological momentary assessment data. *Statistics in Medicine*. 2020;39:577–590. <https://doi.org/10.1002/sim.8429>

APPENDIX

We consider a simple dataset with four variables. In this example, y is the outcome, x_1 is an observation-level covariate, x_2 is a subject-level covariate, and id is the subject identifier.

A.1 | SAMPLE CODE FOR PROC NLMIXED IN SAS

For contrast, we use capital letters for SAS-specific keywords and lowercase letters for variable and parameter names. As shown below, first we provide the PROC name and options (in this case we set one option, which is the number of quadrature points—SAS often uses a single quadrature point by default, which sometimes leads to incorrect estimation). The PARMS statement is used to declare the parameters for NLMIXED to estimate and to set their starting values. (In order to choose reasonable starting values, it may be beneficial to use PROC MIXED to estimate a model without a random scale effect, as it does not require starting values. However, we recommend setting the random scale standard deviation to a value slightly greater than 0, as a value of 0 can lead to convergence issues.) Next, we specify equations for the mean and error variance (*var*), respectively. To improve convergence, we specify the random effects as independent, estimating the Cholesky of the variance/covariance matrix, and then use ESTIMATE statements to produce the estimated variance/covariance values. The MODEL statement specifies the distribution for the outcome variable, and the RANDOM statement defines the random effects, giving the mean vector and variance-covariance matrix (stacked, without symmetric duplicates), as well as the subject grouping variable. The option OUT= instructs SAS to estimate the Empirical Bayes values of the random effects for each subject and place them in the specified dataset. RUN tells SAS to process the preceding statements.

The corresponding coefficients follow the same numbering scheme as the covariates, with 0 representing the intercept. b is used for the regression coefficients (β) and t is used for the scale fixed effects (τ). The three random effects are denoted as u_0 (location intercept), u_1 (location slope), and u_2 (scale). The variance terms for the location random effects are var_0 for the intercept, var_1 for the slope, and var_2 for the scale, with the covariance numbered accordingly.

```

PROC NLMIXED data=test qpoints=11;
PARMS b0=34.83 b2=-7.37 b1=-1.47 T11=6.85 T21=0.69 T22=2.71
t0=4.44 t2=.13 t1=0.08 T33=.005 T31=0 T32=0;
z = (b0+T11*u0) + (b1+T21*u0+T22*u1)*x1 + b2*x2;
vare = EXP(t0 + t1*x1 + t2*x2 + u2*T33 + T31*u0 + T32*u1);
MODEL y ~ NORMAL(z,vare);
RANDOM u0 u1 u2 ~ NORMAL([0,0,0], [1,0,1,0,0,1]) SUBJECT=id OUT=ebestimates;
ESTIMATE "var0" T11*T11;
ESTIMATE "cov01" T11*T21;
ESTIMATE "var1" T21*T21+T22*T22;
ESTIMATE "cov02" T11*T31;
ESTIMATE "cov12" T21*T31+T32*T22;
ESTIMATE "var2" T31*T31+T32*T32+T33*T33;
RUN;

```

A.2 | CODE FOR MERLIN IN STATA

In the code below (based on code provided at <https://www.mjcrowther.co.uk/software/merlin/longitudinal/level1/>), we first provide a statement to install Merlin and then provide code for a linear mixed model using `mixed` and Merlin code for the same model. In Stata, lowercase/uppercase matters, and so most of the code below is in lowercase. However, random effects are designated as uppercase M1 and M2 in Merlin. The id variable is indicated in the brackets so that the random effects are subject-level. Merlin has the possibility of different distributions for the outcome, so `family(gaussian)` is for a normally distributed outcome. The `covariance(unstructured)` specification allows the random effects to be correlated. Then, in order to run the MELS model, one needs to define in `mata` the function `lev1_logl`. For this, we followed the specifications at the Merlin website. Here, `y` denotes the dependent variable, `linpred1` are the mean model predictors, and `varresid` is for the modeling of the WS variance. This function is then called by Merlin to specify and estimate the MELS model. Here, we are indicating that the modeling of the outcome is utilizing the `lev1_logl` `mata` function. Additionally, M1 is for the random location intercept, M2 is for the random location slope, and M3 is for the scale random effect. These are allowed to be correlated by the `covariance(unstructured)` specification. For more information, refer to the website provided.

```

* install Merlin
ssc install merlin

* random intercept and trend model via mixed mixed
y x2 x1 || id: x1, covariance(unstructured) mle nolog

* random intercept and trend model via Merlin
merlin (y x2 x1 M1[id]@1 x1#M2[id]@1, family(gaussian)), ///
       covariance(unstructured) nolog

* defining the function lev1_logl that will be called by Merlin mata:
real matrix lev1_logl(gml)
{
    y = merlin_util_depvar(gml) //response
    linpred1 = merlin_util_xzb(gml) //main lin. pred.
    varresid = exp(merlin_util_xzb_mod(gml,2)) //lev1 lin. pred
    return(lnnormalden(y,linpred1,sqrt(varresid))) //logl
}
end

* MELS model using Merlin model
merlin (y x2 x1 M1[id]@1 x1#M2[id]@1, family(user, llfunction(lev1_logl))) ///
       (M3[id]@1 x2 x1, family(null)) , covariance(unstructured) nolog

```

A.3 | MIXREGLS USE

The original version of MIXREGLS can be obtained at <https://www.jstatsoft.org/article/view/v052i12>. As mentioned, this version only includes one random location effect, in addition to the random scale effect. The manual of this program describes how it can be accessed via R, and a macro program “runmixregls”³⁶ allows access using Stata. This macro program is available here: <https://www.jstatsoft.org/article/view/v059c02>. The extended version of MIXREGLS that includes multiple random location effects and was used for the results presented in this paper can be obtained from the first author of this paper. This extended version is also accessible via R, and work is underway to allow the runmixregls macro program to access this extension as well.

A.4 | TRANSFORMING FROM CHOLESKY TO VARIANCE/COVARIANCE MATRIX

$$\begin{bmatrix} v_{1i} \\ v_{2i} \\ \omega_i \end{bmatrix} = \begin{bmatrix} T_1 & 0 & 0 \\ T_2 & T_3 & 0 \\ \tau^{\theta_1} & \tau^{\theta_2} & \sigma_\omega \end{bmatrix} \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \\ \vartheta_i \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} T_1 & 0 & 0 \\ T_2 & T_3 & 0 \\ \tau^{\theta_1} & \tau^{\theta_2} & \sigma_\omega \end{bmatrix} \begin{bmatrix} T_1 & T_2 & \tau^{\theta_1} \\ 0 & T_3 & \tau^{\theta_2} \\ 0 & 0 & \sigma_\omega \end{bmatrix}$$

$$\sigma_1^2 = T_1^2$$

$$\sigma_{12} = T_1 T_2$$

$$\sigma_2^2 = T_2^2 + T_3^2$$

$$\sigma_{13} = T_1 \tau^{\theta_1}$$

$$\sigma_{23} = T_1 \tau^{\theta_2} + T_3 \tau^{\theta_1}$$

$$\sigma_3^2 = (\tau^{\theta_1})^2 + (\tau^{\theta_2})^2 + (\sigma_\omega)^2$$