RESEARCH ARTICLE

Fast estimation of mixed-effects location-scale regression models

Nathan Gill[1] | Donald Hedeker[2]

[1]Division of Biostatistics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

[2]Department of Public Health Sciences, University of Chicago, Chicago, Illinois, USA

Correspondence
Nathan Gill, Division of Biostatistics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.
Email: nathan.gill@northwestern.edu

Summary
As a result of advances in data collection technology and study design, modern longitudinal datasets can be much larger than they historically have been. Such "intensive" longitudinal datasets are rich enough to allow for detailed modeling of the variance of a response as well as the mean, and a flexible class of models called mixed-effects location-scale (MELS) regression models are commonly used to do so. However, fitting MELS models can pose computational challenges related to the numerical evaluation of multi-dimensional integrals; the slow runtime of current methods is inconvenient for data analysis and makes bootstrap inference impractical. In this paper, we introduce a new fitting technique, called FastRegLS, that is considerably faster than existing techniques while still providing consistent estimators for the model parameters.

KEYWORDS
computational methods, longitudinal data, random effects models

1 | INTRODUCTION

As a result of advances in data collection technology and study design, modern longitudinal datasets can be much larger than they historically have been. In particular, studies involving smartphone-based data collection, wearable sensors, or Ecological Momentary Assessment (EMA) methodology can result in datasets with a very large number of observations (eg, 30+) per subject.[1–3] For example, in longitudinal psychology studies, subjects may be asked to self-report their mood via smartphone or handheld computer many times in a short observation window.[4] In studies of physical activity, researchers may have access to exercise volume data from wearable sensors at regular and frequent intervals,[5–7] potentially resulting in hundreds or thousands of observations per participant. Datasets resulting from studies like these are often called "intensive" longitudinal datasets.[5,6,8]

Intensive longitudinal data are rich enough to allow for detailed modeling of the variance of a response as well as the mean, and there is often scientific interest in doing so.[9,10] In longitudinal studies of mood, for example, a researcher might wish to know if a certain covariate is associated with a more stable (ie, lower variance) mood. Such questions cannot be answered by the traditional models used for longitudinal data, such as the mixed-effects model, which has only a single parameter to model the residual variance. This is not a huge limitation when analyzing smaller datasets, where it would not generally be feasible to estimate additional variance parameters with meaningful precision. However, for intensive longitudinal data, more general models are needed.

The mixed-effects location-scale (MELS) regression model was developed to address this shortcoming of the traditional methods, and is now a common technique for analyzing intensive longitudinal data.[11–14] It has also been used to analyze other types of clustered data, such as survey data[15] and data from education research.[16] The model allows fixed effects and subject-level random effects to influence the mean of a response as well as the variation around that mean. The fixed effects used to model the variance allow researchers to investigate the association between covariates and response stability, while the random effects used to model the variance capture subject-level variability in response stability not captured by the fixed-effects. In a mood study, say, even after controlling for covariates, some subjects might have more, and some others less, stable moods than others, and it is exactly this sort of difference that the random effects capture.

A limitation of the MELS model is that it is difficult to fit computationally. Current techniques compute and maximize the marginal likelihood using Gaussian quadrature to integrate out the random effects, but this can be very slow, especially as the number of random effects increases. It can therefore be inconvenient in settings where many slightly different models are fit and examined, such as exploratory data analysis and model building and checking. Moreover, due to the computational cost, using resampling methods like the bootstrap for inference on the model parameters is not practical, leaving researchers without a robust way to estimate standard errors or determine significance of model parameters. Researchers have also used Bayesian methods to fit the MELS model.[17–20] These techniques are natural given the hierarchical structure of the MELS model, but they are not meant to reduce computational burden. We will restrict ourselves to the frequentist interpretation of the model.

In this paper, we introduce a new fitting technique for the MELS model, called FastRegLS, that is a great deal faster (as much as 100× for some datasets) and more numerically stable than quadrature-based maximum likelihood, while still providing consistent parameter estimates. Because it is so much faster, FastRegLS has the added benefit of enabling bootstrap inference for the MELS model parameters, and can considerably speed up exploratory data analysis and model diagnostics. Such procedures have always been computationally intensive for MELS models, but they are feasible using FastRegLS. We note also that FastRegLS can be used in tandem with existing techniques. For example, one could check how reasonable model-based maximum likelihood Wald intervals are by computing the FastRegLS bootstrap intervals and comparing - a difference could be evidence of deviation from the specified model. In addition, FastRegLS could be used to provide starting values for a maximum likelihood algorithm.

Our method is implemented in an R package, `FastRegLS`, which is included in the supplemental material. However, an additional advantage of FastRegLS is that it can be easily implemented using any software that can fit linear mixed-effects models; unlike quadrature-based methods, sophisticated numerical computing is not required.

The organization of this paper is as follows. In section 2, we give additional background and a formal statement of the MELS model. In section 3, current maximum likelihood-based fitting methods are described. In section 4, we describe our new method, FastRegLS. In section 5, we present empirical verification of the theoretical properties of FastRegLS, and compare FastRegLS with maximum likelihood on both real and simulated data. We conclude in section 6 with a discussion.

2 | THE MELS MODEL

Suppose we have longitudinal data **y** collected from N total subjects, with n_i observations per subject, $i = 1, \dots, N$, for a total of $T = \sum_{i=1}^{N} n_i$ data points. Consider first the basic mixed-effects model commonly fit to such data:

$$y_{ij} = \mathbf{x}_{ij}^{\mathsf{T}}\beta + \nu_i + \epsilon_{ij}. \tag{1}$$

$$\nu_i \sim N(0, \sigma_\nu^2) \tag{2}$$

$$\epsilon_{ij} \sim N(0, \sigma^2) \tag{3}$$

Here y_{ij} is the jth observation from subject i, $i = 1, \dots, N$, which we assume is a Gaussian response variable. The index j runs from 1 to n_i. The vector $\mathbf{x}_{ij} \in \mathbb{R}^k$ is a covariate vector corresponding to the single observation y_{ij}, a single row of a larger design matrix $\mathbf{X} \in \mathbb{R}^{T \times k}$, and $\beta \in \mathbb{R}^k$ is the vector of regression coefficients. The subject-level random effect ν_i follows a $N(0, \sigma_\nu^2)$ distribution, and the residual error ϵ_{ij} follows a $N(0, \sigma^2)$ distribution. The random effect variance σ_ν^2 is

also called the between-subjects (BS) variance, since it characterizes how much subjects differ from the global baseline mean determined by β. The residual variance σ^2 is also called the within-subjects (WS) variance, since it characterizes how much the observations y_{ij} differ from the mean value $\mathbf{x}_{ij}^\mathsf{T}\beta + v_i$. Since σ^2 is a constant, the WS variance is not dependent on subject identity or covariates. The MELS model extends this model by replacing the residual error distribution (3) with the following:

$$\epsilon_{ij} \sim N(0, \sigma_{ij}^2) \tag{4}$$

$$\sigma_{ij}^2 = \exp(\mathbf{w}_{ij}^\mathsf{T}\tau + \tau_l v_i + \omega_i) \tag{5}$$

$$\omega_i \sim N(0, \sigma_\omega^2). \tag{6}$$

The WS variance is now the exponential (because the variance must be positive) of a linear predictor. It is dependent on covariates $\mathbf{W} \in \mathbb{R}^{T \times p}$, of which the $\mathbf{w}_{ij}^\mathsf{T}$ are rows, with a corresponding regression vector $\tau \in \mathbb{R}^p$. Throughout this paper, we will assume that the WS variance model has an intercept term, that is, that the first column of $\mathbf{W}$ is a vector of ones. The third term on the right-hand side of (5) is an additional subject-level random effect ω_i, which also follows a normal distribution. Just as v_i captures subject-level variation in the mean of $\mathbf{y}$ above and beyond the specified covariates $\mathbf{x}_{ij}$, so ω_i captures subject-level variation in the variance of $\mathbf{y}$ above an beyond the covariates $\mathbf{w}_{ij}$. From now on, v_i will be referred to as the random location effect, and ω_i as the random scale effect (giving the MELS model its name). The second term in (5) represents the association between the random location effect and the random scale effect. A similar effect can be achieved by allowing v_i and ω_i to be correlated, but we prefer this regression-like association for its modeling flexibility.[14] Throughout this paper, the term "MELS model" will refer to the model defined by (1), (2), (4)–(6), and its parameters will be denoted by $\Theta = (\beta, \tau, \tau_l, \sigma_v^2, \sigma_\omega^2)$. We note that although we have chosen a longitudinal data setting to explain the MELS model, and will continue to refer to the groupings of correlated observations as "subjects," the same model can be used for clustered data more generally, in which the "subjects" might be geographical locations, schools, and so forth.

3 | MAXIMUM LIKELIHOOD ESTIMATION

Maximum likelihood (ML) estimation remains the most common way to fit the MELS model. It is implemented, for example, in software like MixRegLS,[14] Merlin,[21] and the PROC NLMIXED procedure in SAS.[22,23] The random location and scale effects are replaced by their standardized versions, and Gaussian quadrature is used to compute the necessary integrals. Specifically, if we let $\theta_i^l = \frac{v_i}{\sigma_v}$ and $\theta_i^s = \frac{\omega_i}{\sigma_\omega}$, then θ_i^l and θ_i^s are independent $N(0, 1)$ random variables, and

$$y_{ij}|\theta_i^l, \theta_i^s \sim N\left(\mathbf{x}_{ij}^\mathsf{T}\beta + \sigma_v\theta_i^l, \exp(\mathbf{w}_{ij}^\mathsf{T}\tau + \tau_l\sigma_v\theta_i^l + \sigma_\omega\theta_i^s)\right). \tag{7}$$

Letting $f(y_{ij}|\theta_i^l, \theta_i^s)$ denote the Gaussian likelihood corresponding to (7), the full likelihood conditional on $\Theta^l = (\theta_1^l, \theta_2^l, \dots, \theta_N^l)$ and $\Theta^s = (\theta_1^s, \theta_2^s, \dots, \theta_N^s)$ is given by

$$f(\mathbf{y}|\Theta^l, \Theta^s) = \Pi_{i,j}f(y_{ij}|\theta_i^l, \theta_i^s).$$

The full unconditional likelihood $L(y)$ is then obtained by integrating out Θ^l and Θ^s,

$$L(y) = \int f(\mathbf{y}|\Theta^l, \Theta^s)g(\Theta^l)g(\Theta^s)\mathrm{d}\Theta^l\mathrm{d}\Theta^s, \tag{8}$$

where $g(\cdot)$ is the density of a standard normal random vector in $\mathbb{R}^N$.

No closed form expression exists for (8), so when maximizing $L(y)$, (8) and its derivatives are approximated using Gaussian quadrature: the integral is replaced by a sum in which the integrand is evaluated at a pre-specified grid of points. If there are Q points per dimension, every integral requires Q^2 integrand evaluations per optimization iteration. As the data examples in the Results section demonstrate, this procedure can be time intensive. In addition to being inconvenient for routine data analysis and model building, it makes resampling methods like the bootstrap infeasible. The option

to bootstrap is particularly important for the MELS model and other mixed-effects models, since the variances based on the Fisher information at convergence (eg, as computed in Reference 24) do not always work well for variance components.[25] In the next section, we present FastRegLS, a new technique that provides consistent estimators at a fraction of the computation time of the current methods.

4 | METHODS

4.1 | Illustration on a simpler model

FastRegLS is an extension of an approach of Harvey[26] for estimating the parameters of the simpler heteroskedastic regression model

$$
\begin{aligned}
y_i &= \mathbf{x}_i^\intercal \beta + \epsilon_i \\
\epsilon_i &\sim N(0, \sigma_i^2) \\
\sigma_i^2 &= \exp(\mathbf{w}_i^\intercal \tau),
\end{aligned}
$$

where $w_{i1} = 1$ for all i (ie, the variance model contains an intercept), in which the model is fit in two steps. First, β is estimated using least squares,

$$
\hat{\beta} = (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{y}.
$$

The residuals $\hat{\epsilon}_i = y_i - \mathbf{x}_i^\intercal \hat{\beta}$ are then formed, and Harvey[26] observes that asymptotically,

$$
\log \hat{\epsilon}_i^2 = \mathbf{w}_i^\intercal \tau + \psi_i,
$$

where $\psi_i \sim \log \chi_1^2$. The variance regression parameter τ is then estimated using

$$
\hat{\tau} = (\mathbf{W}^\intercal \mathbf{W})^{-1} \mathbf{W}^\intercal \log \hat{\epsilon}^2 + \mathbb{E} \log \chi_1^2 \mathbf{e}_1, \tag{9}
$$

where $\log \hat{\epsilon}$ denotes the vector whose i^{th} component is $\log \hat{\epsilon}_i$, $\mathbf{e}_1$ denotes the unit vector with 1 in the first position, that is, $\mathbf{e}_1 = (1, 0, \ldots, 0)^\intercal$, and $\mathbb{E} \log \chi_1^2 = 1.27$. It can be shown that (9) is asymptotically unbiased, and has variance

$$
\mathrm{Var}(\hat{\tau}) = \mathrm{Var}(\log \chi_1^2)(\mathbf{W}^\intercal \mathbf{W})^{-1},
$$

where $\mathrm{Var}(\log \chi_1^2) = 4.93$. The estimator (9) is less efficient than the ML estimator, which has variance $2(\mathbf{W}^\intercal \mathbf{W})^{-1}$. However, the ML estimator must be obtained through an iterative optimization algorithm, whereas (9) can be computed in closed form and is much faster.

We now extend this approach to estimate the parameters of the MELS model. The FastRegLS algorithm consists of a parameter update and an initialization. We will describe the updates first, and then describe the initialization. See Algorithm 1 for a full summary of the algorithm. Notice the computational simplicity of the steps, which are either closed-form or can be solved using existing fast algorithms (such as lme4[27] for the random effects models). Note also the mix of EM-like updates and direct likelihood maximizations. For example, the inner-loop updates of $\hat{\beta}$ and the random intercepts $\hat{v}_i$ coincide with their EM analogues. On the other hand, the WS variance parameters and the final out-of-loop $\hat{\beta}$ estimator rely on full likelihood fits of submodels chosen to guarantee good statistical properties (see the Properties section).

4.2 | Parameter updates

Assume that we have estimates $\hat{\Theta}_k = \left(\hat{\beta}_k, \hat{\tau}_k, \hat{\tau}_{l,k}, \hat{\sigma}_{v,k}^2, \hat{\sigma}_{\omega,k}^2 \right)$ from a previous iteration of the algorithm, as well as estimates of the random location and scale effects themselves, $\hat{v}_{i,k}$ and $\hat{\omega}_{i,k}$ for $i = 1, \ldots, N$. Let $\hat{\sigma}_{ij,k+1}^2$ be the plug-in estimate of the

---

Algorithm 1. FastRegLS

1: Fit by maximum likelihood the model

$$y_{ij} = \mathbf{x}_{ij}^\intercal b_0 + \nu_{i,0} + \epsilon_{ij,0}$$
$$\nu_{i,0} \sim N(0, g_0^2)$$
$$\epsilon_{ij,0} \sim N(0, \sigma_0^2)$$

2: $\hat{\beta}_0 \leftarrow \hat{b}_0$
3: $\hat{\sigma}_{\nu,0}^2 \leftarrow \hat{g}_0^2$
4: $\hat{\nu}_{i,0} \leftarrow \left(\frac{1}{\hat{g}_0^2} + \frac{n_i}{\hat{\sigma}_0^2}\right)^{-1} \sum_{j=1}^{n_i} \frac{y_{ij} - \mathbf{x}_{ij}^\intercal \hat{b}_0}{\hat{\sigma}_0^2}$
5: $\delta_{ij,0} \leftarrow \log(y_{ij} - \mathbf{x}_{ij}^\intercal \hat{\beta}_0 - \hat{\nu}_{i,0})^2$
6: Fit by maximum likelihood the model
7:

$$\delta_{ij,0} = \mathbf{w}_{ij}^\intercal t_0 + t_{l,0} \hat{\nu}_{i,0} + \omega_{i,0} + \psi_{ij,0}$$
$$\omega_{i,0} \sim N(0, s_0^2)$$
$$\psi_{ij,0} \sim N(0, \kappa_0^2)$$

8: $\hat{\tau}_0 \leftarrow \hat{t}_0 + \mathbb{E} \log \chi_1^2$
9: $\hat{\tau}_{l,0} \leftarrow \hat{t}_{l,0}$
10: $\hat{\sigma}_{\omega,0}^2 \leftarrow \hat{s}_0^2$
11: $\hat{\omega}_{i,0} \leftarrow \left(\frac{1}{s_0^2} + \frac{n_i}{\kappa_0^2}\right)^{-1} \sum_{j=1}^{n_i} \frac{\delta_{ij,0} - \mathbf{w}_{ij}^\intercal \hat{t}_0 - \hat{t}_{l,0} \hat{\nu}_{i,0}}{\hat{\kappa}_0^2}$
12: **for** $k = 0, 1, 2 \ldots$ **do**
13:      $\hat{\sigma}_{ij,k+1}^2 \leftarrow \exp(\mathbf{w}_{ij}^\intercal \hat{\tau}_k + \hat{\tau}_{l,k} \hat{\nu}_{i,k} + \hat{\omega}_{i,k})$
14:      $\hat{\nu}_{i,k+1} \leftarrow \left(\frac{1}{\hat{\sigma}_{\nu,k}^2} + \sum_{j=1}^{n_i} \frac{1}{\hat{\sigma}_{ij,k+1}^2}\right)^{-1} \sum_{j=1}^{n_i} \frac{y_{ij} - \mathbf{x}_{ij}^\intercal \hat{\beta}_k}{\hat{\sigma}_{ij,k+1}^2}$
15:      $V_{k+1} \leftarrow (\underbrace{\hat{\nu}_{1,k+1}, \ldots, \hat{\nu}_{1,k+1}}_{n_1 \text{ times}}, \underbrace{\hat{\nu}_{2,k+1}, \ldots, \hat{\nu}_{2,k+1}}_{n_2 \text{ times}}, \ldots, \underbrace{\hat{\nu}_{N,k+1}, \ldots, \hat{\nu}_{N,k+1}}_{n_N \text{ times}})^\intercal$
16:      $\mathbf{W}_{k+1} \leftarrow$ the diagonal matrix whose diagonal is the vector

$$(\underbrace{\hat{\sigma}_{11,k+1}^{-2}, \ldots, \hat{\sigma}_{1n_1,k+1}^{-2}}_{n_1 \text{ entries}}, \underbrace{\hat{\sigma}_{21,k+1}^{-2}, \ldots, \hat{\sigma}_{2n_2,k+1}^{-2}}_{n_2 \text{ entries}}, \ldots, \underbrace{\hat{\sigma}_{N1,k+1}^{-2}, \ldots, \hat{\sigma}_{Nn_N,k+1}^{-2}}_{n_N \text{ entries}})^\intercal$$

17:      $\hat{\beta}_{k+1} \leftarrow \left(\mathbf{X}^\intercal \mathbf{W}_{k+1} \mathbf{X}\right)^{-1} \mathbf{X}^\intercal \mathbf{W}_{k+1} (\mathbf{y} - V_{k+1})$
18:      $\delta_{ij,k+1} \leftarrow \log(y_{ij} - \mathbf{x}_{ij}^\intercal \hat{\beta}_{k+1} - \hat{\nu}_{i,k+1})^2$
19:      Fit by maximum likelihood the model

$$\delta_{ij,k+1} = \mathbf{w}_{ij}^\intercal t_{k+1} + t_{l,k+1} \hat{\nu}_{i,k+1} + \omega_{i,k+1} + \psi_{ij,k+1}$$
$$\omega_{i,k+1} \sim N(0, s_{k+1}^2)$$
$$\psi_{ij,k+1} \sim N(0, \kappa_{k+1}^2)$$

20:      $\hat{\tau}_{k+1} \leftarrow \hat{t}_{k+1} + \mathbb{E} \log \chi_1^2 \mathbf{e}_1$
21:      $\hat{\tau}_{l,k+1} \leftarrow \hat{t}_{l,k+1}$
22:      $\hat{\sigma}_{\omega,k+1}^2 \leftarrow \hat{s}_{k+1}^2$
23:      $\hat{\omega}_{i,k+1} \leftarrow \left(\frac{1}{\hat{s}_{k+1}^2} + \frac{n_i}{\hat{\kappa}_{k+1}^2}\right)^{-1} \sum_{j=1}^{n_i} \frac{\delta_{ij,k+1} - \mathbf{w}_{ij}^\intercal \hat{t}_{k+1} - \hat{t}_{l,k+1} \hat{\nu}_{i,k+1}}{\hat{\kappa}_{k+1}^2}$
24:      $\hat{\sigma}_{\nu,k+1}^2 \leftarrow \frac{1}{N} \sum_{i=1}^{N} \hat{\nu}_{i,k+1}^2$
25: **end for**
26: Fit by maximum likelihood the model

$$y_{ij} = \mathbf{x}_{ij}^\intercal b_{k_{\max}} + \nu_{i,k_{\max}} + \epsilon_{ij,k_{\max}}$$
$$\nu_{i,k_{\max}} \sim N(0, g_{k_{\max}}^2)$$
$$\epsilon_{ij,k_{\max}} \sim N(0, \exp(\mathbf{w}_{ij}^\intercal \hat{\tau}_{k_{\max}} + \hat{\tau}_{l,k_{\max}} \hat{\nu}_{i,k_{\max}} + \hat{\omega}_{i,k_{\max}}))$$

27: $\hat{\beta}_{k_{\max}} \leftarrow \hat{b}_{k_{\max}}$
28: $\hat{\sigma}_{\nu,k_{\max}}^2 \leftarrow \hat{g}_{k_{\max}}^2$

WS variance based on $\hat{\Theta}_k$:

$$\hat{\sigma}^2_{ij,k+1} = \exp\left(\mathbf{w}^\intercal_{ij}\hat{\tau}_k + \hat{\tau}_{l,k}\hat{v}_{i,k} + \hat{\omega}_{i,k}\right).$$

We update the estimates of the random location effects by computing their expected value conditional on the observed data $\mathbf{y}$ and the current parameter estimates:

$$\hat{v}_{i,k+1} = \mathbb{E}(v_i|\mathbf{y}, \hat{\Theta}_k, \hat{\omega}_{i,k})$$
$$= \left(\frac{1}{\hat{\sigma}^2_{v,k}} + \sum_{j=1}^{n_i}\frac{1}{\hat{\sigma}^2_{ij,k+1}}\right)^{-1}\sum_{j=1}^{n_i}\frac{y_{ij} - \mathbf{x}^\intercal_{ij}\hat{\beta}_k}{\hat{\sigma}^2_{ij,k+1}}. \qquad (10)$$

Note that this is essentially an empirical Bayes procedure, where the hyperparameter of the prior (2) is determined using the data.

Next, we update our estimate of β using weighted least squares: $y_{ij} - \hat{v}_{i,k+1}$ is regressed on $\mathbf{X}$, using the $\hat{\sigma}^2_{ij,k+1}$ as weights, and $\hat{\beta}_{k+1}$ is the resulting regression vector. More precisely, let $V_{k+1} \in \mathbb{R}^T$ be the vector containing each $\hat{v}_{i,k+1}$ repeated n_i times,

$$V_{k+1} = \left(\underbrace{\hat{v}_{1,k+1}, \ldots, \hat{v}_{1,k+1}}_{n_1 \text{ times}}, \underbrace{\hat{v}_{2,k+1}, \ldots, \hat{v}_{2,k+1}}_{n_2 \text{ times}}, \ldots, \underbrace{\hat{v}_{N,k+1}, \ldots, \hat{v}_{N,k+1}}_{n_N \text{ times}}\right)^\intercal,$$

and let $\mathbf{W}_{k+1} \in \mathbb{R}^{T\times T}$ be the diagonal matrix whose diagonal is the vector

$$\left(\underbrace{\hat{\sigma}^{-2}_{11,k+1}, \ldots, \hat{\sigma}^{-2}_{1n_1,k+1}}_{n_1 \text{ entries}}, \underbrace{\hat{\sigma}^{-2}_{21,k+1}, \ldots, \hat{\sigma}^{-2}_{2n_2,k+1}}_{n_2 \text{ entries}}, \ldots, \underbrace{\hat{\sigma}^{-2}_{N1,k+1}, \ldots, \hat{\sigma}^{-2}_{Nn_N,k+1}}_{n_N \text{ entries}}\right)^\intercal.$$

The updated β is then given by

$$\hat{\beta}_{k+1} = (\mathbf{X}^\intercal\mathbf{W}_{k+1}\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{W}_{k+1}(\mathbf{y} - V_{k+1}).$$

We update β in this way to avoid fitting a second (see below for the first) mixed-effects model within the loop, which would slow down the computation time. However, we do update β a final time out of the loop using a mixed-effects model, see below.

Next, following Harvey[26] to update the WS variance parameters, we form the residuals

$$\hat{\epsilon}_{ij,k+1} = y_{ij} - \mathbf{x}^\intercal_{ij}\hat{\beta}_{k+1} - \hat{v}_{i,k+1},$$

and, letting $\delta_{ij,k+1} = \log\hat{\epsilon}^2_{ij,k+1}$, fit the mixed-effects model

$$\delta_{ij,k+1} = \mathbf{w}^\intercal_{ij}t_{k+1} + t_{l,k+1}\hat{v}_{i,k+1} + \omega_{i,k+1} + \psi_{ij,k+1}$$
$$\omega_{i,k+1} \sim N(0, s^2_{k+1})$$
$$\psi_{ij,k+1} \sim N(0, \kappa^2_{k+1})$$

using ML, obtaining parameter estimates $\hat{t}_{k+1}, \hat{t}_{l,k+1}, \hat{s}^2_{k+1}$, and $\hat{\kappa}^2_{k+1}$. We then set

$$\hat{\tau}_{k+1} = \hat{t}_{k+1} + \mathbb{E}\log\chi^2_1\mathbf{e}_1$$
$$\hat{\tau}_{l,k+1} = \hat{t}_{l,k+1}$$
$$\hat{\sigma}^2_{\omega,k+1} = \hat{s}^2_{k+1}$$
$$\hat{\omega}_{i,k+1} = \left(\frac{1}{\hat{s}^2_{k+1}} + \frac{n_i}{\hat{\kappa}^2_{k+1}}\right)^{-1}\sum_{j=1}^{n_i}\frac{\delta_{ij,k+1} - \mathbf{w}^\intercal_{ij}\hat{t}_{k+1} - \hat{t}_{l,k+1}\hat{v}_{i,k+1}}{\hat{\kappa}^2_{k+1}}, \qquad (11)$$

where (11) is analogous to (10), the earlier update of the random location effects. Finally, we update σ_v^2 using the sample variance of the $\hat{v}_{i,k+1}$,

$$\hat{\sigma}_{v,k+1}^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{v}_{i,k+1}^2.$$

Finally, once we have finished iterating in this way for $k_{\max}$ iterations, we obtain final estimates of β and σ_v^2 by fitting the mixed-effects model,

$$y_{ij} = \mathbf{x}_{ij}^\intercal b_{k_{\max}} + v_{i,k_{\max}} + \epsilon_{ij,k_{\max}}$$
$$v_{i,k_{\max}} \sim N\left(0, g_{k_{\max}}^2\right)$$
$$\epsilon_{ij,k_{\max}} \sim N\left(0, \hat{\sigma}_{ij,k_{\max}}^2\right).$$

Our final parameter estimates are then $\hat{\beta} = \hat{b}_{k_{\max}}$, $\hat{\tau} = \hat{\tau}_{k_{\max}}$, $\hat{\tau}_l = \hat{\tau}_{l,k_{\max}}$, $\hat{\sigma}_v^2 = \hat{g}_{k_{\max}}^2$, and $\hat{\sigma}_\omega^2 = \hat{\sigma}_{\omega,k_{\max}}^2$, collectively denoted $\hat{\Theta} = (\hat{\beta}, \hat{\tau}, \hat{\tau}_l, \hat{\sigma}_v^2, \hat{\sigma}_\omega^2)$.

4.3 | Initialization

The initialization is similar to the in-loop updates. We begin by fitting the homoskedastic mixed-effects model

$$y_{ij} = \mathbf{x}_{ij}^\intercal b_0 + v_{i,0} + \epsilon_{ij,0} \tag{12}$$

$$v_{i,0} \sim N\left(0, g_0^2\right)$$
$$\epsilon_{ij,0} \sim N\left(0, \sigma_0^2\right) \tag{13}$$

using ML, and set $\hat{\beta}_0 = \hat{b}_0$ and $\hat{\sigma}_{v,0}^2 = \hat{g}_0^2$. As in the in-loop updates, we initialize the random location effects by their expected values under (12)–(13) conditional on the observed data **y** and our parameter estimates:

$$\hat{v}_{i,0} = \mathbb{E}(v_i | \mathbf{y}, \hat{b}_0, \hat{g}_0^2, \hat{\sigma}_0^2)$$
$$= \left(\frac{1}{\hat{g}_0^2} + \frac{n_i}{\hat{\sigma}_0^2}\right)^{-1} \sum_{j=1}^{n_i} \frac{y_{ij} - \mathbf{x}_{ij}^\intercal \hat{b}_0}{\hat{\sigma}_0^2}.$$

To initialize the WS variance parameters, letting

$$\delta_{ij,0} = \log\left(y_{ij} - \mathbf{x}_{ij}^\intercal \hat{\beta}_0 - \hat{v}_{i,0}\right)^2,$$

we fit the model

$$\delta_{ij,0} = \mathbf{w}_{ij}^\intercal t_0 + t_{l,0} \hat{v}_{i,0} + \omega_{i,0} + \psi_{ij,0}$$
$$\omega_{i,0} \sim N(0, s_0^2)$$
$$\psi_{ij,0} \sim N(0, \kappa_0^2)$$

using ML, and set $\hat{\tau}_0 = \hat{t}_0 + \mathbb{E} \log \chi_1^2$, $\hat{\tau}_{l,0} = \hat{t}_{l,0}$, and $\hat{\sigma}_{\omega,0}^2 = \hat{s}_0^2$. Finally, we initialize the random scale effects following (11):

$$\hat{\omega}_{i,0} = \left(\frac{1}{\hat{s}_0^2} + \frac{n_i}{\hat{\kappa}_0^2}\right)^{-1} \sum_{j=1}^{n_i} \frac{\delta_{ij,0} - \mathbf{w}_{ij}^\intercal \hat{t}_0 - \hat{t}_{l,0} \hat{v}_{i,0}}{\hat{\kappa}_0^2}.$$

4.4 | Properties

Theorem 1 (Consistency). *The FastRegLS estimators are consistent for their respective parameters (regardless of k_{max}).*

The estimators $\hat{\Theta}$ resulting from the FastRegLS algorithm are consistent for their respective parameters (see the supplementary material for a proof, and the Results section for an empirical demonstration). Importantly, the estimators remain consistent no matter how many times we execute the inner loop, that is, how large or small k_{max} is. The results presented in the Results section are all obtained using $k_{max} = 1$, except where noted. As k_{max} increases, the FastRegLS estimates tend to "converge" in the sense that the difference between successive estimates grows small, but this is not necessary for valid estimation. In practice, it is a good idea to verify that k_{max} is large enough that the change in the estimates from iteration to iteration is small. Such large changes are possible when there are significant deviations from the specified model in the data used for fitting. For example, one of our simulations compares the performance of bootstrap to Wald confidence intervals using a dataset with some of the residual errors drawn from an exponential distribution (Table 3). To ensure that the FastRegLS estimates were stable, we used $k_{max} = 5$ for this simulation after observing large changes in the estimates from the first 2–3 iterations. We note also that increasing k_{max} does not necessarily produce estimates that are closer to those obtained using ML, although both converge to the true parameter values, and therefore get closer to each other, as the sample size increases.

There are two methods available to obtain standard errors for the FastRegLS estimators. The first is a Wald-type interval based on approximations to the asymptotic covariances of the estimators. It is difficult to derive exact formulas for these, but approximations that work well in practice can be obtained under the assumption that the parameter estimates used in the intermediate steps are known exactly (eg, treating the $\hat{v}_i$ as fixed, known covariates when estimating the WS variance parameters). Formulas and details of the derivation are provided in the supplementary material.

Though the Wald intervals are convenient and work well in practice (see the Results section for simulations demonstrating good coverage rates), they (1) condition on variable quantities and therefore will tend to underestimate the uncertainties of the estimators. and (2) are model-dependent. The impact of (1) can be seen in Figure 2: although they still perform well, the intervals for WS variance parameters, which depend most on approximations of this type, tend to have lower coverage rates than the intervals for the other parameters. The nonparametric bootstrap is an alternative that accounts for all sources of variation, and is also more robust to model misspecification. The ability to bootstrap is a major advantage of FastRegLS over existing MELS model fitting techniques, which are not fast enough to perform a large number of iterations in a practical amount of time.

5 | RESULTS

In this section, we present several applications of FastRegLS and compare its performance with that of ML. We also compare the performance of bootstrap and Wald intervals for FastRegLS. Unless otherwise specified, all results were obtained using $k_{max} = 1$.

5.1 | Simulations

5.1.1 | Consistency

To empirically demonstrate the consistency of the FastRegLS estimators $\hat{\Theta}$, we simulated data from the MELS model with $\beta \in \mathbb{R}^4$, $\tau \in \mathbb{R}^4$, each randomly generated with $N(0, 1)$ entries, $\tau_l = 0.3$, $\sigma_v^2 = 1$, and $\sigma_\omega^2 = 1$, for a variety of sample sizes. Figure 1 shows the mean squared error (MSE) of $\hat{\Theta}$ as a function of both the number of subjects in the sample N and the number of observations per subject n_i (each subject had the same number of observations). At each combination of N and n_i, 100 datasets were generated using the above parameter values, and the design matrices X and W had random $N(0, 1)$ entries (except an intercept column of ones). The plotted point is the average MSE over the 100 runs when estimating the parameters using FastRegLS, and the the error bars are plus/minus 2 standard deviations of this average. (Specifically, the MSE for each iteration was calculated as $\frac{1}{11}||\Theta - \hat{\Theta}||_2^2$, since $\theta \in \mathbb{R}^{11}$ for this simulation, and the plotted MSE is the average of this quantity over all iterations.) The MSE approaches 0 as the sample size increases, but notice
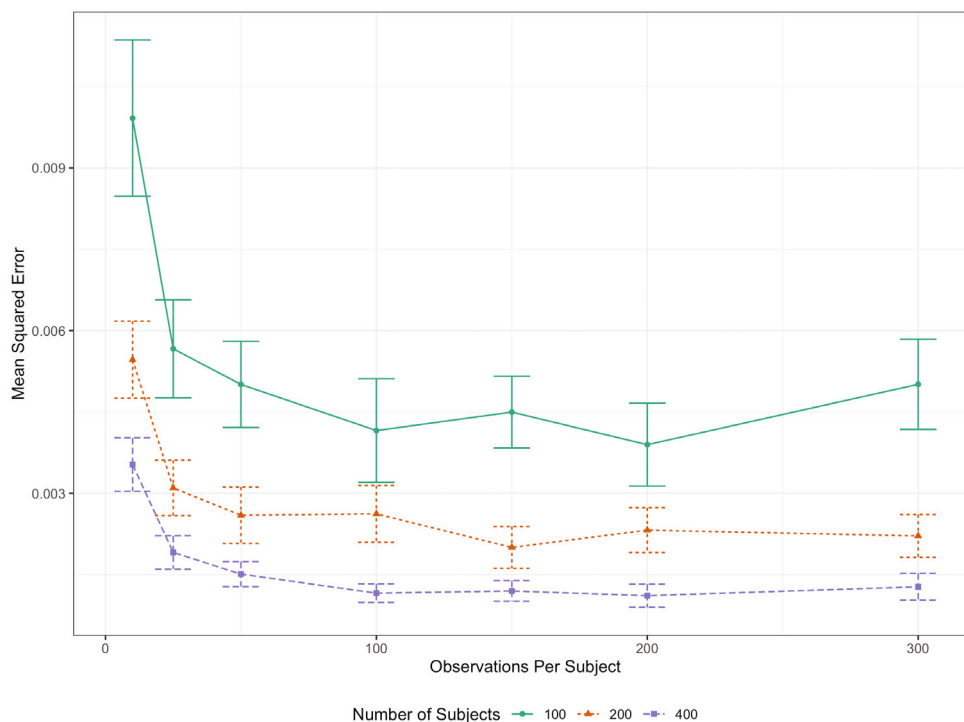
FIGURE 1 Mean squared error (MSE) of FastRegLS estimator $\hat{\Theta}$ as a function of both the number of subjects in the sample N and the number of observations per subject n_i (each subject had the same number of observations). At each combination of N and n_i, 100 datasets were generated using the parameter values in Table 2, and the design matrices X and W had random $N(0, 1)$ entries (except an intercept column of ones). The plotted point is the average MSE over the 100 runs, and the the error bars are plus/minus 2 standard deviations of this average. The MSE approaches 0 as the sample size increases

TABLE 1 Comparison of runtimes (in seconds) between FastRegLS and maximum likelihood implemented with the MixRegLS software. FastRegLS is orders of magnitude faster, and the difference increases as the datasets increase in size.

Method	20	30	40	50	60	70	80	90	100
FastRegLS	0.38	0.46	0.51	0.79	0.64	0.80	0.72	0.81	0.88
MixRegLS	19.78	29.83	40.34	47.62	64.57	73.79	84.61	101.13	106.80

Note: All datasets had $N = 200$ subjects, with the same parameter values as in Table 2, and n_i increasing from 20 to 100.

that we need both $N \to \infty$ and $n_i \to \infty$. This is because N limits how precisely we can estimate the variance components σ_v^2 and σ_ω^2, and thus for fixed N the MSE trajectory plateaus above 0.

5.1.2 | Run time

Because it does not require numerical integration in two dimensions, FastRegLS is much faster than existing techniques. Table 1 compares the run times (in seconds) of FastRegLS with those of MixRegLS[14] on a sequence of increasingly large datasets. All datasets had $N = 200$ subjects, with the same parameter values as in Table 2, and n_i increasing from 20 to 100. FastRegLS is orders of magnitude faster, and the difference increases as the datasets increase in size.

5.1.3 | Performance of confidence intervals

In order to test the performance of the Wald and bootstrap FastRegLS confidence intervals, we simulated 100 datasets from a MELS model with $\beta \in \mathbb{R}^4$, $\tau \in \mathbb{R}^4$, each randomly generated with $N(0, 1)$ entries, $\tau_l = 0.3$, $\sigma_v^2 = 1$, $\sigma_\omega^2 = 1$, $N = 200$,

TABLE 2 For 100 datasets generated using the parameter values indicated in column 2, and 200 subjects with 30 observations per subject, we fit a MELS model using FastRegLS and constructed both types of 95% confidence intervals, Wald intervals and bootstrap intervals. The coverage columns shows the proportion of times each type of interval contained the true parameter value, as well as the average widths of the intervals over the 100 runs. The observed coverage rates are close to the target of 0.95, and widths of the two types of intervals are similar. Note that since this is simulated data and there is no model misspecification, the Wald intervals tend to slightly outperform the bootstrap intervals.

Parameter	Value	Boot. coverage	Wald coverage	Boot. avg. width	Wald avg. width
β_0	−0.626	0.97	0.98	0.266	0.285
β_1	0.184	0.90	0.91	0.033	0.034
β_2	−0.836	0.94	0.93	0.035	0.035
β_3	1.595	0.90	0.93	0.036	0.036
τ_0	0.330	0.97	0.97	0.299	0.306
τ_1	−0.820	0.88	0.91	0.117	0.116
τ_2	0.487	0.91	0.94	0.121	0.121
τ_3	0.738	0.91	0.89	0.119	0.117
τ_l	0.300	0.90	0.96	0.295	0.316
σ_v^2	1.000	0.92	0.96	0.594	0.411
σ_ω^2	1.000	0.94	0.96	0.438	0.466

TABLE 3 Comparison of the coverage probabilities of 95 % bootstrap and Wald confidence intervals for a dataset with a fraction of the residual errors drawn from an exponential distribution (ie, under model misspecification)

Type	β_0	β_1	β_2	β_3	τ_0	τ_1	τ_2	τ_3	τ_l	σ_v^2	σ_ω^2
Wald	0.96	0.95	0.96	0.95	0.19	0.56	0.85	0.60	0.98	0.94	0.89
Bootstrap	0.93	0.93	0.96	0.92	0.64	0.82	0.96	0.81	0.98	0.95	0.96

Note: The bootstrap intervals outperform the Wald intervals, but would be very time consuming to compute using the traditional ML fitting techniques for MELS models.

$n_i = 30$ for $i = 1, \dots, N$, and the design matrices X and W with random $N(0, 1)$ entries (except an intercept column of ones). For each such dataset, we fit the model using FastRegLS and constructed both types of 95% confidence intervals. Table 2 shows the proportion of times the interval contained the true parameter value, as well as the average widths of the intervals over the 100 runs. The observed coverage rates are close to the target of 0.95, and the widths of the two types of intervals are similar.

To demonstrate the need for the bootstrap when we suspect model deviations, we repeated the coverage probability simulation with a mixture of exponential and Gaussian residual error terms. With probability 0.8, the residual error term was drawn from the usual $N(0, \sigma_{ij}^2)$ distribution, but with probability 0.2, the residual error was drawn from $\sigma_{ij}(\text{Exp}(1) - 1)$, an exponential distribution shifted to have mean 0 and scaled to have variance σ_{ij}^2. Table 3 compares the coverage rates of the 95% Wald and bootstrap intervals in this setting. Because the model deviations produced large changes in the FastRegLS estimates for the first few iterations, we used $k_{\max} = 5$ for this table. The Wald intervals perform poorly on the WS variance parameters, but we can obtain more reasonable coverage rates using the bootstrap (note, however, that both perform poorly for τ_0). The bootstrap intervals also perform better for σ_ω^2. Without FastRegLS, such intervals could not be computed in a practical amount of time.

We also tested the Wald interval coverage rates on datasets of varying sizes. As in the consistency simulation, we fixed the number of subjects at $N = 200$, and increased the number of observations per subject from 20 to 100. For each sample size, we simulated 500 datasets using the same parameter values as in Table 2, fit a MELS model to each dataset using FastRegLS, and for each parameter, determined the proportion of times the 95% Wald confidence interval contained the true value. The results are shown in Figure 2. For all parameters and for all sample sizes, the coverage rates stay close to the target of 0.95.
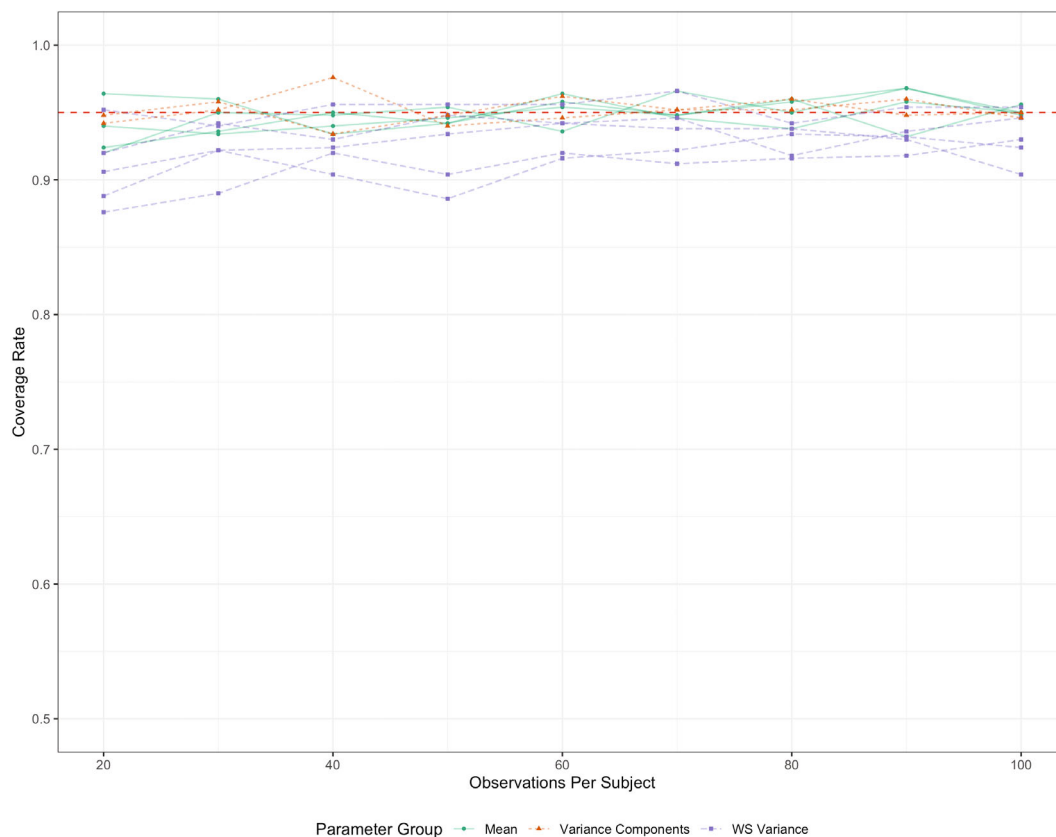
FIGURE 2 Additional coverage rate simulations for the Wald confidence intervals with $N = 200$ subjects and varying number of observations per subject (*x*-axis). Each plotted point is the coverage rate for a particular MELS model parameter. The points are colored based on parameter grouping for visualization purposes. For all parameters and sample sizes, the coverage rates stay close to the target of 0.95

5.1.4 | Comparison with ML on simulated data

To directly compare the performance of FastRegLS with ML, we simulated data from a MELS model using the parameter values in Table 4, and then used both FastRegLS and ML (run using MixRegLS) to obtain parameter estimates and confidence intervals. The results are shown in Table 4. For each method, the point estimates are close to the true values, and both types of 95% confidence interval (bootstrap and Wald) contain the true value, but FastRegLS obtains these results at a fraction of the computation time (Table 1).

5.2 | Real data

5.2.1 | Comparison with ML on real data

We also compared the performance of FastRegLS and ML on real data from an intensive longitudinal study of first-year university students.[28] The study followed 82 first-year university students for 46 days, during which time the students took their final exams. Data was collected using daily online surveys designed to capture various aspects of lifestyle, including mood, physical activity, sleep patterns, and so forth. Only data from the first 32 days of the study are publicly available, and this is the dataset we analyze. There were 2111 total responses, with 29.3 responses per student on average (and a range of 8 to 23).

We will focus on positive affect (PA), an average of the extent to which participants felt several emotions (happy, content, cheerful, sad, downhearted, frustrated) on a given day, measured using a 1–7 Likert scale. Psychology researchers are often interested in the variability of a measure in addition to its mean value, and because of this PA data is often analyzed using MELS models.[12,13,17] We can confirm that this is reasonable if we examine plots of the PA trajectories

TABLE 4 Comparison of parameter estimates and confidence intervals from FastRegLS and from maximum likelihood (run using MixRegLS) on simulated data

Parameter	Value	FastReg	FastReg Wald	FastReg Boot.	ML	ML wald	ML Boot.
β_0	−0.62	−0.63	(−0.77, −0.49)	(−0.78, −0.49)	−0.60	(−0.74, −0.45)	(−0.75, −0.46)
β_1	0.18	0.18	(0.16, 0.20)	(0.16, 0.20)	0.19	(0.17, 0.20)	(0.17, 0.20)
β_2	−0.84	−0.83	(−0.85, −0.82)	(−0.85, −0.82)	−0.83	(−0.87, −0.79)	(−0.85, −0.811)
β_3	1.60	1.59	(1.57, 1.61)	(1.57, 1.61)	1.59	(1.57, 1.61)	(1.57, 1.62)
τ_0	0.33	0.39	(0.23, 0.55)	(0.22, 0.57)	0.29	(0.14, 0.44)	(0.15, 0.44)
τ_1	−0.82	−0.87	(−0.93, −0.81)	(−0.95, −0.82)	−0.83	(−0.85, −0.81)	(−0.87, −0.79)
τ_2	0.49	0.49	(0.44, 0.55)	(0.44, 0.56)	0.49	(0.45, 0.53)	(0.45, 0.53)
τ_3	0.74	0.74	(0.68, 0.80)	(0.68, 0.81)	0.73	(0.69, 0.77)	(0.68, 0.76)
τ_l	0.30	0.35	(0.18, 0.51)	(0.18, 0.53)	0.26	(0.11, 0.41)	(0.12, 0.40)
σ_v^2	1.00	0.95	(0.75, 1.16)	(0.67, 1.39)	1.01	(0.82, 1.24)	(0.78, 1.22)
σ_ω^2	1.00	1.13	(0.87, 1.39)	(0.86, 1.46)	0.92	(0.74, 1.13)	(0.85, 1.07)

Note: For each method, point estimates are close to the true values, and both types of 95% confidence interval (bootstrap and Wald) contain the true value, but FastRegLS obtains these results at a fraction of the computation time (Table 1).

of several study subjects (Figure 3). Just by looking, we can see that these trajectories differ greatly in their mean and variance; by fitting a MELS model, we can determine whether such differences are due to covariates, or to individual-level heterogeneity captured by random location and scale effects.

We fit a MELS model with PA as response, and age and sex of the student as both mean and WS variance covariates. Specifically, the model fit was

$$\text{PA}_{ij} = \beta_0 + \beta_{\text{age}}\text{age}_i + \beta_{\text{sex}}\text{sex}_i + v_i + \epsilon_{ij}$$
$$v_i \sim N(0, \sigma_v^2)$$
$$\epsilon_{ij} \sim N(0, \sigma_{ij}^2)$$
$$\sigma_{ij}^2 = \exp(\tau_0 + \tau_{\text{age}}\text{age}_i + \tau_{\text{sex}}\text{sex}_i + \tau_l v_i + \omega_i)$$
$$\omega_i \sim N(0, \sigma_\omega^2),$$

where PA_{ij} is the positive affect measurement of student i on day j, age_i is the age of student i in years, and sex_i is the sex of student i (0 = Female, 1 = Male). Table 5 shows the point estimates and confidence intervals from FastRegLS and ML (the latter run using MixRegLS). The point estimates of the regression coefficients are very similar for the two methods, as are both types of confidence intervals. The most interesting difference between the two is the random scale variance, for which the FastRegLS point estimate is larger than that of ML. However, we note that the confidence intervals do still overlap. The key point is that the interpretation of the coefficients is the same for each method. For instance, both methods tell us that males have a much smaller variability in their PA responses ($\tau_{\text{sex}} < 0$), and that higher average PA responses (better moods) are associated with smaller PA variability ($\tau_l < 0$). There is also subject-level variability beyond age and sex that is captured by the random location and scale effects ($\sigma_v^2 > 0$ and $\sigma_\omega^2 > 0$). However, the FastRegLS estimates come at a fraction of the computation time: 0.19 s, compared to 8.25 s for ML.

6 | DISCUSSION

In this paper, we introduced a new fitting technique, called FastRegLS, for the MELS model that is considerably faster than currently available maximum likelihood-based techniques. The FastRegLS estimators are consistent, and approximations to their asymptotic variances are available. Because it is so much faster than existing techniques, FastRegLS makes bootstrap inference practical for the MELS model for the first time. Given the complexity of the MELS model and the challenges associated with performing inference on variance components, this is a useful addition to the MELS model toolkit.
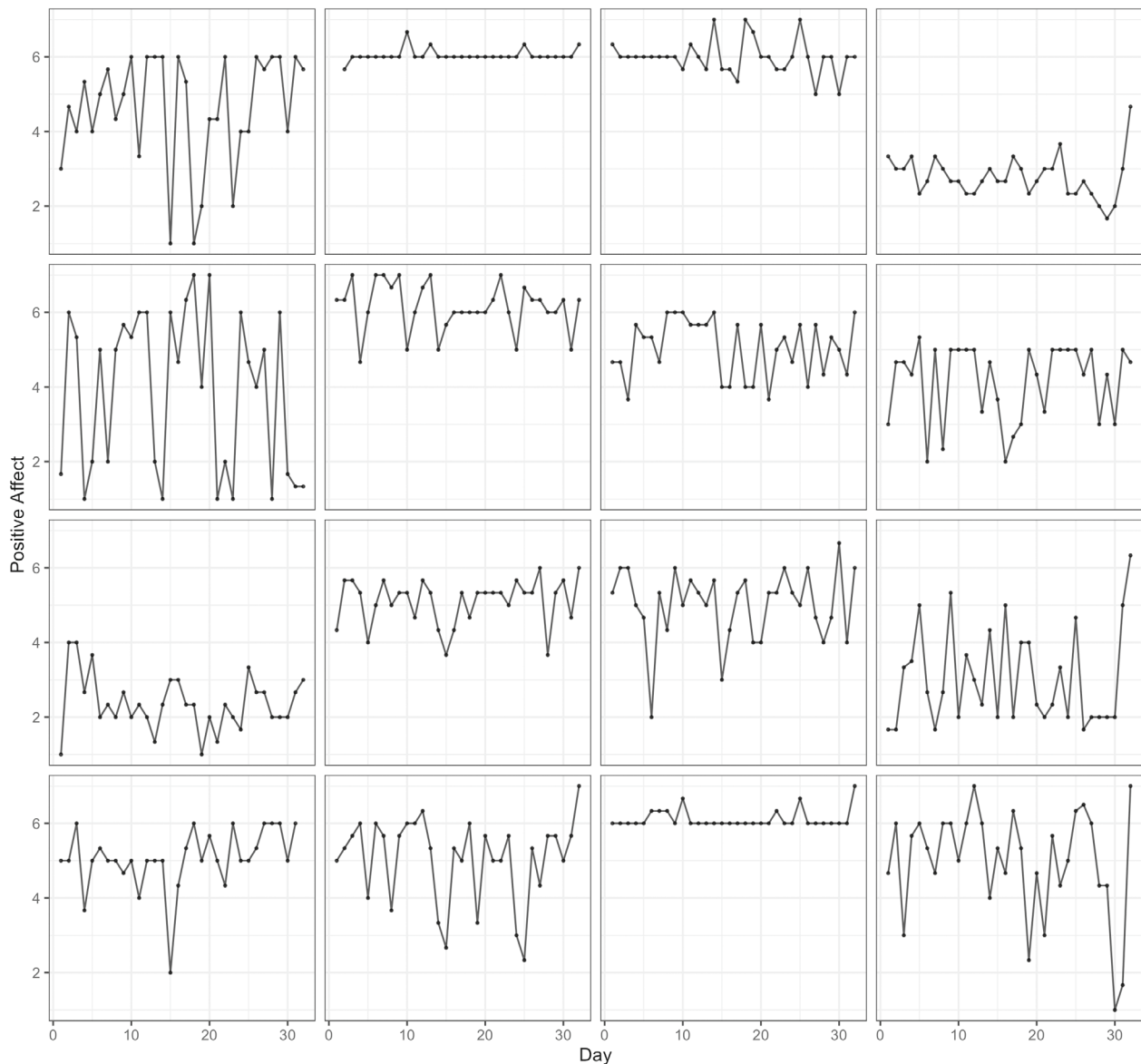
FIGURE 3 Positive affect (PA) trajectories for a selection of subjects in our example intensive longitudinal study. The trajectories differ considerably in mean and variance; by fitting a MELS model, we can determine whether such differences are due to covariates, or to individual-level heterogeneity captured by random location and scale effects

We emphasize that FastRegLS does not need to replace a full ML solution in order to be useful. Even if a researcher plans to eventually fit a model using ML, FastRegLS could still be useful by speeding up exploratory data analysis. Or if one suspects deviations from the specified model, they could check how reasonable ML Wald confidence intervals are by computing the FastRegLS bootstrap intervals and comparing - if they are similar, then there is more reason to trust the ML results. FastRegLS could also be used to quickly provide starting values for a full ML algorithm.

Although we only considered a MELS model with a random location intercept and a random scale intercept, recent work[29] has considered additional location and scale random effects, for example random slopes. For these models, the existing approaches are even more computationally burdensome, since the complexity of the fitting problem increases exponentially in the number of random effects. The FastRegLS approach of breaking down the full model into submodels could potentially provide even more substantial speed increases in this setting, since the complexity is limited only by the size of the largest of these submodels. Future work will extend FastRegLS to allow for more complicated random effects structures. Additionally, since the models considered in this paper are only appropriate for Gaussian responses, we plan to extend FastRegLS to models for non-Gaussian response variables.

TABLE 5 Comparison of parameter estimates and confidence intervals from FastRegLS and from maximum likelihood (run using MixRegLS) on real data from a longitudinal study on positive affect

Parameter	FastReg	FastReg wald	FastReg boot	ML	ML wald	ML boot
β_0	4.00	(2.79, 5.21)	(2.42, 5.24)	3.98	(1.97, 5.02)	(1.87, 5.28)
β_{age}	0.01	(−0.04, 0.06)	(−0.05, 0.08)	0.01	(−0.04, 0.06)	(−0.05, 0.11)
β_{sex}	0.49	(−0.38, 1.35)	(−0.42, 1.28)	0.49	(−0.37, 1.35)	(−0.42, 1.25)
τ_0	0.52	(−0.52, 1.55)	(−0.42, 2.82)	0.71	(−0.15, 1.58)	(−0.17, 2.72)
τ_{age}	−0.02	(−0.07, 0.02)	(−0.13, 0.02)	−0.03	(−0.07, 0.01)	(−0.13, 0.01)
τ_{sex}	−0.89	(−1.63, −0.15)	(−2.12, −0.04)	−0.75	(−1.37, −0.13)	(−1.56, −0.02)
τ_l	−0.34	(−0.57, −0.11)	(−0.62, −0.09)	−0.32	(−0.52, −0.11)	(−0.51, −0.04)
σ_v^2	1.27	(1.06, 1.47)	(0.38, 1.76)	1.26	(0.87, 1.82)	(0.82, 1.62)
σ_ω^2	0.81	(0.62, 1.01)	(0.36, 1.17)	0.50	(0.31, 0.73)	(0.48, 0.82)

Note: The point estimates and confidence intervals are generally similar for the two methods. However, the FastRegLS runtime is much faster than ML - 0.19 s versus 8.25 s.

DATA AVAILABILITY STATEMENT
The positive affect dataset is publicly available at https://doi.org/10.7910/DVN/27388.

ORCID
Nathan Gill https://orcid.org/0000-0002-0815-785X

REFERENCES
1. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. *J Happiness Stud: Interdiscip Forum Subj Well-Being*. 2003;4:35-52.
2. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Ann Rev Clin Psychol*. 2008;4:1-32.
3. Barrett LF, Barrett DJ. An introduction to computerized experience sampling in psychology. *Soc Sci Comput Rev*. 2001;19:175-185.
4. Mermelstein R, Hedeker D, Flay B, Shiffman S. Situational versus intra-individual contributions to adolescents' subjective mood experience of smoking. In *Annual Meeting for the Society for Research on Nicotine and Tobacco*. Savannah, Georgia; 2002.
5. Phillips SM, Welch WA, Fanning J, et al. Daily physical activity and symptom reporting in breast cancer patients undergoing chemotherapy: an intensive longitudinal examination. *Cancer Epidemiol Biomarkers Prev*. 2020;29:2608-2616.
6. Auster-Gussman LA, Gavin KL, Siddique J, et al. Social cognitive variables and physical activity during chemotherapy for breast cancer: an intensive longitudinal examination. *Psychooncol*. 2021;31:425-435.
7. Seshadri DR, Li RT, Voos JE, et al. Wearable sensors for monitoring the physiological and biochemical profile of the athlete. *NPJ Digi Med*. 2019;2(72).
8. Walls TA, Schafer JL, eds. *Models for Intensive Longitudinal Data*. Oxford: Oxford University Press; 2006.
9. Carroll R. Variances are not always nuisance parameters. *Biometrics*. 2003;59(2):211-220.
10. Davidian M, Carroll R. Variance function estimation. *J Am Stat Assoc*. 1987;82(400):1079-1091.
11. Cleveland WS, Denby L, Liu C. *Random Scale Effects*. Tech. Rep., Murray Hill: Bell Labs; 2002.
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(2):627-634.
13. Hedeker D, Mermelstein RJ, Demirtas H. Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Stat Med*. 2012;31(27):3328-3336.
14. Hedeker D, Nordgren R. MIXREGLS: a program for mixed-effects location scale analysis. *J Stat Softw*. 2013;52(12):107-120.
15. Brunton-Smith I, Sturgis P, Leckie G. Detecting, and understanding interviewer effects on survey data using a cross-classified mixed-effects location scale model. *J R Stat Soc, Ser A*. 2017;180:551-568.
16. 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.
17. 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;47:177-200.
18. Lin X, Mermelstein RJ, Hedeker D. A 3-level bayesian mixed effects location scale model with an application to ecological momentary assessment data. *Stat Med*. 2018;37(13):2108-2119.
19. Williams DR, Martin SR, Liu S, Rast P. Bayesian multivariate mixed-effects location scale modeling of longitudinal relations among affective traits, states, and physical activity. *Eur J Psychol Assess*. 2020;36(6):981-997.

20. Blozis SA. Bayesian two-part multilevel model for longitudinal media use data. *J Mark Anal*. 2022;10;311-328. doi:10.1057/s41270-022-00172-9

21. Crowther MJ. Merlin-A unified modeling framework for data analysis and methods development in stata. *Stat J*. 2020;20:763-784.

22. SAS Institute. *SAS User's Guide: Statistics*. Vol 2. Cary: Sas Inst; 1985.

23. Ma Q, Hedeker D. Modeling of between- and within-subject variances using mixed effects location scale (MELS) models. *SAS 2020 Glob Forum Proc*. 2020;4181.

24. Wand M. Fisher information for generalised linear mixed models. *J Multivar Anal*. 2007;98:1412-1416.

25. Faraway JJ. *Extending the Linear Model with R*. Boca Raton: CRC Press; 2016.

26. Harvey AC. Estimating regression models with multiplicative heteroskedasticity. *Econom*. 1976;44(3):461-465.

27. Bates D, Machler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67:1-48.

28. Flueckiger L, Lieb R, Meyer AH, Mata J. How health behaviors relate to academic performance via affect: an intensive longitudinal study. *PLOS One*. 2014;9(10). doi:10.7910/DVN/27388

29. Nordgren R, Hedeker D, Dunton G, Yang CH. Extending the mixed-effects model to consider within-subject variance for ecological momentary assessment data. *Stat Med*. 2020;39(5):577-590.

SUPPORTING INFORMATION

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