

Estimation of the linear mixed integrated Ornstein–Uhlenbeck model

Rachael A. Hughes^a, Michael G. Kenward^b, Jonathan A. C. Sterne^a and Kate Tilling^a

^aSchool of Social and Community Medicine, University of Bristol, Bristol, UK; ^bDepartment of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

ABSTRACT

The linear mixed model with an added integrated Ornstein–Uhlenbeck (IOU) process (linear mixed IOU model) allows for serial correlation and estimation of the degree of derivative tracking. It is rarely used, partly due to the lack of available software. We implemented the linear mixed IOU model in Stata and using simulations we assessed the feasibility of fitting the model by restricted maximum likelihood when applied to balanced and unbalanced data. We compared different (1) optimization algorithms, (2) parameterizations of the IOU process, (3) data structures and (4) random-effects structures. Fitting the model was practical and feasible when applied to large and moderately sized balanced datasets (20,000 and 500 observations), and large unbalanced datasets with (non-informative) dropout and intermittent missingness. Analysis of a real dataset showed that the linear mixed IOU model was a better fit to the data than the standard linear mixed model (i.e. independent within-subject errors with constant variance).

ARTICLE HISTORY

Received 26 April 2015
Accepted 25 December 2016

KEYWORDS



Fixed effects; Newton Raphson; Integrated Ornstein–Uhlenbeck process; random effects; repeated measures

AMS SUBJECT CLASSIFICATION

62F99; 62J99; 62M10; 62P10

1. Introduction

Linear mixed models, proposed by Laird and Ware [1], are commonly used for the analysis of longitudinal clinical markers ('biomarkers') of disease; for example HIV research on the time course of CD4 counts [2] and time course of progesterone in a menstrual cycle [3]. In such settings the data are typically unbalanced: the number of measurements differs among subjects and the time interval between consecutive measurements differs within and between subjects. The variance of the biomarker may be nonstationary (vary over time) and, when measurements on the same individual are recorded close together in time, within-subject measurements may be serially correlated (also known as autocorrelation). Typically, the linear mixed model is applied assuming no within-subject serial correlation ('conditional independence'). However, the conditional independence model can allow for nonstationary variance through between-subject random effects: for example, a model with a random intercept and a random linear slope implies that the variance is a quadratic function of time. The stochastic linear mixed model is the conditional independence model with an added Gaussian stochastic process to model within-subject serial correlation. Diggle [4] proposed a stochastic linear mixed model where the covariance structure was assumed to be stationary and modelled by a random intercept, independent measurement errors and a stationary Gaussian stochastic process (the Gaussian correlation function). Taylor et al. [5], Zhang et al. [6] and Stirrup et al. [7] allowed for a nonstationary covariance structure by including in their model random effects other than the random intercept and a nonstationary stochastic process (the integrated Ornstein–Uhlenbeck (IOU))

CONTACT Rachael A. Hughes  rachael.hughes@bristol.ac.uk  School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

process, the nonhomogeneous Ornstein–Uhlenbeck process and the fractional Brownian Motion process respectively). This paper focuses on the model proposed by Taylor et al., which we shall refer to as the linear mixed IOU model.

Based on model fit statistics (Akaike Information Criterion and log-likelihood) Taylor et al. [5] and Wolbers et al. [8] concluded that, compared to a conditional independence model, the linear mixed IOU model gave an improved fit to longitudinal CD4 counts. Based on a simulation study, Taylor and Law [9] concluded that, when predicting future measurements in subjects, a linear mixed IOU model was more robust than a conditional independence model (with a random linear or quadratic slope) to incorrect specification of the true covariance structure of the data. However the linear mixed IOU model is rarely used in practice, due to the lack of available software.

The purpose of this paper is to describe the estimation and implementation of the linear mixed IOU model within Stata, and to evaluate, using simulations, the feasibility and practicality of estimating the linear mixed IOU model by restricted maximum likelihood (REML) estimation. We compare REML estimation of the model under different: (1) parameterizations of the IOU stochastic process, (2) longitudinal data structures, (3) optimization algorithms and (4) random effects structures. We also evaluate our implementation when applied to unbalanced data typically seen in large cohort studies. In Section 2 we describe the linear mixed IOU model, the estimation procedure and our implementation in Stata 13.1 [10]. Sections 3 and 4 present the methods and results of our simulation study respectively. In Section 5 we apply the linear mixed IOU model to height measurements from Christ's Hospital Study and we conclude with a general discussion in Section 6.

2. Computational methods and theory

2.1. Linear mixed IOU model

The linear mixed IOU model implies that a subject's biomarker trajectory is a realization of an IOU process. Taylor et al. [5] hypothesized that in the short term a subject's biomarker process may maintain the same trajectory but over long periods of time the biomarker's trajectory may vary. A special feature of the IOU process is that it allows for differing levels of derivative tracking; that is, the degree to which a subject's biomarker measurements maintain the same trajectory over long periods of time. The linear mixed IOU model estimates the degree of derivative tracking from the data as opposed to assuming a fixed level via a particular covariance function. Two limiting cases of the model are (1) a specific conditional independence model (strong derivative tracking) and (2) a stochastic linear mixed model with a Brownian Motion process (no derivative tracking).

Consider a dataset of m subjects, where subject i has n_i repeated measurements recorded at time-points $\mathbf{t}_i = \{t_{ij}\}$ ($i = 1, \dots, m; j = 1, \dots, n_i$). For subject i , $\mathbf{y}_i = \{y_{ij}\}$ is a n_i -vector of responses, $\mathbf{X}_i = \{X_{ij}\}$ is a $n_i \times p$ design matrix for the p -vector of fixed effects $\boldsymbol{\beta}$ (i.e. population regression coefficients), \mathbf{Z}_i is a $n_i \times q$ design matrix for the q -vector of random effects \mathbf{u}_i (i.e. subject-specific regression coefficients), $\mathbf{w}_i = \{w_{ij}\}$ is a n_i -vector of realized values of the IOU stochastic process and \mathbf{e}_i is a n_i -vector of independent measurement errors. The random effects \mathbf{u}_i , IOU process \mathbf{w}_i and measurement errors \mathbf{e}_i are assumed to be mutually independent. We assume independence of responses y_i between subjects.

The linear mixed IOU model can be written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \mathbf{w}_i + \mathbf{e}_i,$$

where \mathbf{u}_i , \mathbf{w}_i and \mathbf{e}_i are independently, normally distributed with zero means and covariances \mathbf{G} , \mathbf{H}_i and $\sigma^2\mathbf{I}_{n_i}$, respectively. \mathbf{G} is unstructured and \mathbf{H}_i is defined as follows, for $j_1 = 1, \dots, n_i$, the variance of \mathbf{w}_i at time-point t_{ij_1} is

$$H_i^{j_1j_1} = \frac{\tau^2}{\alpha^3}[\alpha t_{ij_1} + \exp(-\alpha t_{ij_1}) - 1],$$

and, for $j_1, j_2 = 1, \dots, n_i$ and $j_1 \neq j_2$, the covariance of \mathbf{w}_i between time-points t_{ij_1} and t_{ij_2} is

$$H_i^{j_1 j_2} = \frac{\tau^2}{2\alpha^3} [2\alpha \min(t_{ij_1}, t_{ij_2}) + \exp(-\alpha t_{ij_1}) + \exp(-\alpha t_{ij_2}) - 1 - \exp(-\alpha |t_{ij_1} - t_{ij_2}|)].$$

The covariance matrix of \mathbf{y}_i is $\mathbf{V}_i = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i^T + \mathbf{H}_i + \sigma^2 \mathbf{I}_{n_i}$ [11] and we denote the vector of unknown variance parameters by $\boldsymbol{\theta}$, which consists of the unique components of \mathbf{G} , α , τ and σ^2 . When a linear mixed IOU model only contains a single random effect (a random intercept) we shall refer to the model as a ‘random intercept IOU’ model. Similarly, we shall refer to a linear mixed IOU model with a random intercept and random slope as a ‘random slope IOU’ model.

The IOU stochastic process is parameterized by α and τ . According to Taylor et al. [5] the α parameter can be interpreted as a measure of the degree of derivative tracking, where a small value of α indicates a strong degree of derivative tracking (i.e. a subject’s measurements closely follows the same trajectory over long periods). The parameter τ serves as a scaling parameter.

Taylor et al. [5] note that there are two special limiting cases of the linear mixed IOU model. Firstly, as α tends to ∞ (derivative tracking becomes weaker) with τ^2/α^2 held constant \mathbf{w}_i becomes a realization of a Brownian Motion process (also known as the Wiener stochastic process). Taylor and Law [9] called this model the Brownian Motion mixed-effects model. Secondly, for a random intercept IOU model, as α tends to 0 (derivative tracking becomes stronger) with τ^2/α^2 held constant, the model reduces to a conditional independence model with a random intercept and slope, and with zero correlation between the intercept and the slope.

The IOU process is nonstationary because the covariance between w_{ij_1} and w_{ij_2} depends upon time-points t_{ij_1} and t_{ij_2} as well as their difference. Due to nonstationarity the covariance matrix of the random intercept IOU model would be the same if it was applied to (1) a balanced dataset of measurements recorded at t intervals where derivative tracking is α and τ and (2) to a dataset of measurements recorded at Qt intervals where derivative tracking is $Q^{-1}\alpha$ and $Q^{-3/2}\tau$. A disadvantage of nonstationarity is that it requires a definition of a natural time 0, which may not be feasible in some applications. In the absence of a known time 0, Taylor et al. [5] proposed modelling differences; for example, selecting a first measurement y_{iE_i} and modelling the differences $y_{ij} - y_{iE_i}$ for $j > E_i$. However, this also requires the selection of a ‘first measurement’ which is comparable, with respect to the time course of the disease, for all subjects. Note that, a conditional independence model with random intercept and slope (and an unstructured matrix \mathbf{G}) is also nonstationary.

2.2. Estimation

Maximum likelihood (ML) estimators of $\boldsymbol{\theta}$ are biased downwards because the estimation process does not take into account the degrees of freedom used in the estimation of the fixed effects $\boldsymbol{\beta}$ [1,12–14]. For this reason Patterson and Thompson [11] introduced the restricted (or residual) maximum likelihood, a marginal likelihood that is a function of a linearly independent set of residual contrasts and is free of $\boldsymbol{\beta}$. For both REML and ML estimation data can be missing at random [15].

2.2.1. Optimization algorithms

Calculation of REML estimates requires an optimization algorithm such as the Newton Raphson (NR) algorithm [16], Fisher Scoring (FS) algorithm [17], Average Information (AI) algorithm [18] or Expectation Maximization (EM) algorithm [19]. Only the EM algorithm is numerically stable in that the log-likelihood is guaranteed to increase at each iteration and will always converge to estimates within the parameter space. However, the EM algorithm does not give standard errors for the parameter estimates and is slow to converge compared to the NR, FS and AI algorithms [20]. Modified EM algorithms have been proposed to improve the time to convergence; for example using conditional maximization [21–23] and parameter expansion [24, 25].

The FS and AI algorithms are variants of the NR algorithm. The FS algorithm replaces the observed information matrix by the expected information matrix in the NR algorithm, and the AI algorithm replaces the observed information matrix by the average of the observed and expected information matrices (called the AI matrix). The convergence time for the NR algorithm is quicker than the FS algorithm because the NR algorithm converges in fewer iterations and its cost-per-iteration is only slightly slower than that of the FS algorithm [20]. However, the FS algorithm is more robust to poor starting values than the NR algorithm and so Jennrich and Sampson [17] recommended starting with a few iterations of the FS algorithm and then switching to the NR algorithm.

The AI algorithm was developed in the context of animal breeding applications, where estimation can involve manipulation of large matrices. The observed and expected information matrices (used in the NR and FS algorithms respectively) require evaluation of the trace of a matrix. The AI matrix is calculated by taking the sum of squares of expected values, which is quicker to calculate than the trace of a matrix. Gilmour et al. [18] compared the AI and FS algorithms by estimating a linear mixed model when applied to data from a spring wheat trial. The authors concluded that the AI algorithm ‘had similar convergence properties, yet avoided the computing burdens of that approach’ [18].

2.2.2. Implementation of the estimation procedure

Following Pinheiro and Bates [26] we conducted REML estimation with respect to the parameters of the log-Cholesky parameterization of \mathbf{G} . Taylor et al. [5] parameterized the IOU process as α and $\omega = \tau^2/\alpha^2$ and experienced convergence problems as α became increasingly large or small. Taylor et al. suggested (but did not explore) reparameterizing α as $\ln \alpha$ or as α^{-2} if α was suspected to be large. We consider six different parameterizations of the IOU process (see Section 3.1.2, comparison 1): $[\alpha, \tau]$, $[\alpha, \omega]$, $[\ln \alpha, \tau]$, $[\ln \alpha, \omega]$, $[\alpha^{-2}, \tau]$ and $[\alpha^{-2}, \omega]$.

The covariance matrix of \mathbf{y}_i can be expressed as $\mathbf{V}_i = \sigma^2 \mathbf{V}_i^* = \sigma^2 (\mathbf{Z}_i \mathbf{G}^* \mathbf{Z}_i^T + \mathbf{H}_i^* + \mathbf{I}_{n_i})$, where $\mathbf{G}^* = \sigma^{-2} \mathbf{G}$ and $\mathbf{H}_i^* = \sigma^{-2} \mathbf{H}_i$. Therefore, we profile the restricted log-likelihood with respect to σ^2 and so reduce the number of parameters to be optimized. We denote the reduced vector of unknown parameters by $\boldsymbol{\theta}^*$, which consists of the unique elements of $\tilde{\mathbf{G}}^*$ (log-Cholesky parameterization of \mathbf{G}^*), α (or $\ln \alpha$ or α^{-2}) and τ/σ (or ω/σ^2). We estimate $\boldsymbol{\theta}^*$ such that it minimizes the negative of twice the profiled restricted log-likelihood for $\boldsymbol{\theta}^*$. Ignoring the constant terms (and given statistical independence among subjects) the contribution to this function for subject i is

$$-2l_R(\boldsymbol{\theta}^* | \mathbf{y}_i, \mathbf{X}_i) = \ln |\mathbf{V}_i^*| + \mathbf{r}_i^T (\mathbf{V}_i^*)^{-1} \mathbf{r}_i + \ln |\mathbf{X}_i^T (\mathbf{V}_i^*)^{-1} \mathbf{X}_i|,$$

where $\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}} = (\sum_{i=1}^m \mathbf{X}_i (\mathbf{V}_i^*)^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^m \mathbf{X}_i (\mathbf{V}_i^*)^{-1} \mathbf{y}_i)$ is the estimate of $\boldsymbol{\beta}$ given the latest estimate of $\boldsymbol{\theta}^*$.

Once minimization with respect to $\boldsymbol{\theta}^*$ is completed, we transform these parameters such that their ranges are $(-\infty, \infty)$. $(\boldsymbol{\theta}^*, \sigma^2)$ is reparameterized to $\check{\boldsymbol{\theta}} = (\check{g}, \ln \alpha, \ln \omega, \ln \sigma)$, where \check{g} is a vector of the unique elements of $\check{\mathbf{G}}$, which is a reparameterization of \mathbf{G} such that the diagonal elements are the logarithms of standard deviations (i.e. $\check{G}[j_1, j_1] = \ln \sqrt{G[j_1, j_1]}$) and the off-diagonal elements are hyperbolic arctangents of the correlations (i.e. $\check{G}[j_1, j_2] = \tanh^{-1} G[j_1, j_2] \div (\sqrt{G[j_1, j_1]} \sqrt{G[j_2, j_2]})$ for $j_1 \neq j_2$). We obtain the information matrix with respect to $\check{\boldsymbol{\theta}}$, calculate 95% confidence intervals and then transform the endpoints to the required scale (i.e. \mathbf{G} , α , ω and σ).

The optimization algorithms require evaluation of the REML log-likelihood and its first and second partial derivatives. When a dataset contains a large number of measurements computation of these quantities can be prohibitively computer-intensive due to the manipulation and inversion of large matrices. As we assume statistical independence between subjects we computed these quantities separately on each subject, which requires the manipulation of (dimensionally) smaller matrices and so improves computational efficiency. To further improve efficiency, we computed these quantities using the method of Wolfinger et al. [27], which avoids direct computation of \mathbf{V}_i^{-1} by using an extension of the W-transformation [28,29] and the sweep operator [30].

We programmed the estimation procedure in Stata 13.1 using its matrix programming language Mata [31]. We used the inbuilt Mata *optimize* function to perform the optimization.

The optimization algorithm requires starting values for θ^* . We derived the starting values from the data by calculating, across subjects, variances of the response at measurement time-points and covariances of the response between measurement time-points. For example, for the random intercept IOU model the slope of the variances over time is a reasonable approximation for ω . We used the same method for balanced and unbalanced longitudinal data. However, for unbalanced data we first ‘regularized’ the dataset. For example, if measurements were intended to be recorded every 3 months we rounded the ‘observed’ time points to the nearest multiple of 3 months.

3. Simulation study design

The aim of the simulation study was to evaluate our Stata implementation of the linear mixed IOU model. The design of the simulation study was based on Taylor et al.’s [5] analysis of the Los Angeles seroconverters cohort of the Multicentre AIDS Cohort Study. The study recruited homosexual and bisexual men in Los Angeles between 1984 and 1985. The cohort consisted of subjects with a known date of seroconversion (i.e. change from HIV antibody negative to HIV antibody positive) and so the natural time zero (for the time course of the HIV biomarkers) was the time of seroconversion. The biomarker of interest was CD4 count (cells/mm³) and all CD4 counts were transformed by a fourth-root power to approximate a constant within-subject variance. In the notation of Section 2.1 $y_{ij} = \sqrt[4]{\text{CD4 count at time } j \text{ in subject } i}$. Taylor et al.’s analysis includes all observations before the introduction of AZT treatments. Therefore, CD4 counts tend to decrease with increasing time since seroconversion.

We conducted separate simulation studies based on balanced and unbalanced longitudinal data.

3.1. Balanced longitudinal data

The simulation study based on balanced longitudinal data evaluated estimation of the linear mixed IOU parameters under different: (1) parameterizations of the IOU process, (2) data structures, (3) optimization algorithms and (4) random slope IOU models.

3.1.1. Data model

Consider a dataset with m subjects, each with n responses (CD4 counts transformed by a fourth-root power) observed at regular k monthly intervals (i.e. response y_{ij} was observed $j \times k$ months since time of seroconversion). For subject i the n responses $\mathbf{y}_i = (y_{i1}, \dots, y_{in})^T$ were generated under a linear mixed IOU model with a linear population slope

$$\mathbf{y}_i = \beta_0 + \beta_1 \mathbf{t}_i + u_{0i} + u_{1i} \mathbf{t}_i + \mathbf{w}_i + \mathbf{e}_i,$$

where β_0 and u_{i0} respectively denote the fixed and random intercept, β_1 and u_{i1} respectively denote the fixed and random slope, and \mathbf{w}_i and \mathbf{e}_i respectively denote realizations of the IOU and measurement error processes. For all settings $\beta_0 = 5.195$, $\beta_1 = -0.222$ and $\sigma = 0.234$, where $\mathbf{e}_i \sim N(0, \sigma^2)$. For comparisons (1) parameterizations of the IOU process, (2) data structures and (3) optimization algorithms, data were generated under a random intercept IOU model (i.e. $u_{1i} = 0$ for all subjects) and the variance of the random intercept u_{0i} was set to 0.1156. The values of the random effects covariance parameters of the random slope IOU models are discussed in Section 3.1.2 (*Comparison 4*). For all comparisons we repeated the simulations for three levels of derivative tracking: $\alpha = 1.31$ and $\omega = 0.1$ (strong derivative tracking), $\alpha = 3$ and $\omega = 0.5$ (moderate derivative tracking), and $\alpha = 20.7$ and $\omega = 1$ (weak derivative tracking). These values were based on analyses reported in [5, 9].

3.1.2. Comparisons

For all comparisons we generated 1000 datasets and to each simulated dataset we fitted the same linear mixed IOU model that was used to generate the data.

Comparison 1: Parameterization of the IOU process. We compared six different parameterizations of the IOU process: $[\alpha \tau]$, $[\alpha \omega]$, $[\ln \alpha \tau]$, $[\ln \alpha \omega]$, $[\alpha^{-2} \tau]$ and $[\alpha^{-2} \omega]$. We generated data (under a random intercept IOU model) for $m = 1000$ subjects each with $n = 20$ measurements recorded every $k = 3$ months. The values of m, k and n are typical of large HIV cohort studies. Estimation was performed using the NR algorithm.

Comparison 2: Data structures. We considered how the estimation process was influenced by: 2(a) the number of subjects versus the number of measurements per subject (whilst keeping the time interval between consecutive measurements the same), and 2(b) the time interval between consecutive measurements (whilst keeping the number of subjects and the number of measurements per subject constant).

2(a): For a large dataset of 5000 observations in total we compared the following two data structures: $m = 1000$ subjects, $n = 5$ measurements, $k = 3$ months and $m = 250$ subjects, $n = 20$ measurements, $k = 3$ months. Also, for a moderately sized dataset of 500 observations in total we compared data structures: $m = 100$ subjects, $n = 5$ measurements, $k = 3$ months and $m = 25$ subjects, $n = 20$ measurements, $k = 3$ months.

2(b): As discussed at the end of Section 2.1, the degree of derivative tracking depends upon the time interval k . By choosing values for α and ω we were able to change the time interval k whilst keeping the degree of derivative tracking the same across comparisons. We compared the following three settings (which result in the same IOU covariance matrix): monthly intervals with $\alpha = 20.7$ and $\omega = 1$, 3-monthly intervals with $\alpha = 6.9$ and $\omega = \frac{1}{3}$ and yearly intervals with $\alpha = 1.725$ and $\omega = \frac{1}{12}$. We repeated the simulations for a large dataset of 1000 subjects each with 20 measurements and for a moderately sized dataset of 100 subjects each with 5 measurements.

For both data structure comparisons, estimation was performed using the NR algorithm and the optimal IOU parameterization based on the results from comparison 1.

Comparison 3: Optimization algorithms. We compared three optimization algorithms: the NR algorithm, a combination of c iterations of the AI algorithm followed by the NR algorithm, and a combination of c iterations of the FS algorithm followed by the NR algorithm. We repeated the simulations for $c = 3$ and $c = 10$. We compared the algorithms on two different data structures: a large dataset of $m = 20,000$ subjects each with $n = 20$ measurements recorded every $k = 3$ months and on a moderately sized dataset of $m = 100$ subjects each with $n = 5$ measurements recorded every $k = 3$ months, where starting values (calculated from the data) may be further from the truth for the smaller dataset. Estimation was performed using the optimal IOU parameterization based on the results from comparison 1.

Comparison 4: Random slope IOU models. We investigated the conditions under which it was feasible to estimate a model with a random intercept, random slope effect and the IOU process. We compared three random slope IOU models with different values for the random effects covariance matrix. Model RS1: $\text{Var}[u_{0i}] = 0.12$, $\text{Cov}[u_{0i}, u_{1i}] = -0.010$, $\text{Var}[u_{1i}] = 0.010$ (smallest variances and covariance), model RS2: $\text{Var}[u_{0i}] = 0.5$, $\text{Cov}[u_{0i}, u_{1i}] = -0.25$, $\text{Var}[u_{1i}] = 0.25$ and model RS3: $\text{Var}[u_{0i}] = 1.0$, $\text{Cov}[u_{0i}, u_{1i}] = -0.50$, $\text{Var}[u_{1i}] = 0.50$ (largest variances and covariance). Estimation was performed using the NR algorithm and the optimal IOU parameterization based on the results from comparison 1.

3.2. Unbalanced longitudinal data design

We evaluated our implementation of the linear mixed IOU model when applied to unbalanced data. We allowed the number of measurements per subject and the times of the measurements to vary between subjects. Also, the time intervals between consecutive measurements were allowed to vary

between subjects and within a subject. The features of the unbalanced data were based on a UK HIV cohort [32].

3.2.1. Data model

We simulated data on 1489 subjects with 20,000 observations in total: 150 subjects with 2 measurements, 224 subjects with 5 measurements, 372 subjects with 10 measurements and 743 subjects with 20 measurements. We considered two scenarios regarding the time intervals between consecutive measurements.

Scenario 1: Target visits: All subjects were intended to be measured every 3 months. For subject i , the time intervals between consecutive measurements were modelled in days as $N(90, 30^2)$ so that 95% of the time intervals would be between 1 and 5 months duration. All time intervals were required to be at least 7 days.

Scenario 2: Intermittent dropout: Data were simulated allowing for long gaps between consecutive measurements thus generating data with intermittent dropout. For those subjects with 2 or 5 measurements the time intervals between consecutive measurements were modelled as $365 \text{ days} + 90 \times \text{std-lognormal}$, where std-lognormal denotes the standardized form of the lognormal distribution $\exp\{N(0, 0.5^2)\}$. So, 95% of the time intervals would be between 6 and 18 months long. We simulated under the positively skewed lognormal distribution to avoid negative time intervals. For those subjects with 10 or 20 measurements the time intervals were modelled in days as $N(90, 30^2)$. All time intervals were required to be at least 7 days. Under both scenarios, the responses y_i of subject i were generated under a random intercept IOU model with a linear population slope using the same parameters values as discussed in Section 3.1.1.

3.3. Evaluation criteria

We used the same evaluation criteria for the simulation studies on balanced and unbalanced longitudinal data. We classified the estimation process as converged if the optimization algorithm converged to a solution within 100 iterations and the information matrix with respect to θ (used to calculate standard errors) was positive definite (i.e. all eigenvalues of the matrix were greater than 1×10^{-8}). Among the datasets where the estimation process converged we calculated the median and interquartile range (IQR) of the number of iterations to reach convergence. For each variance parameter ν we evaluated: (i) bias of the estimates, (ii) empirical standard deviation of the estimates of its transformation $\check{\nu}$, (iii) median and IQR of the estimated standard errors of $\check{\nu}$ and (iv) coverage percentage of the 95% confidence interval for ν . Note, we were unable to use the Delta method to approximate the standard errors of θ from $\check{\theta}$ because for some values of $\check{\theta}$ the matrix of first derivatives of the transformation function had zero entries, resulting in zero standard errors. Based on 1000 simulations, the Monte Carlo standard error for the true coverage percentage of 95% is $\sqrt{95(100 - 95)/1000} = 0.689$ (3 d.p.) [33], implying that the estimated coverage percentage (to nearest whole percent) should lie within the range 94% and 96% (with 95% probability).

4. Simulation results

4.1. Balanced longitudinal data

In this section we present the results for the comparisons of the parameterizations of the IOU process and the random effects structures when derivative tracking was weak ($\alpha = 20.7$ and $\omega = 1$). The results for the remaining comparisons (structures of the data and optimization algorithms), and moderate and strong derivative tracking are summarized in the text below. All results not reported are available upon request from the authors.

Across all comparisons, the majority of the fixed effects (linear slope and constant) were unbiased. For those biased estimates the magnitude of the bias was very small; for example, 0.0011 (MCSE

0.0005). Note that, theory tells us that fixed effects REML (and ML) estimators from multivariate normal models are unbiased even when the covariance matrix is misspecified [34]. The confidence intervals for the fixed effects had coverage levels between 93% and 96%.

4.1.1. Parameterizations of the IOU process

Table 1 presents the results of estimation under different parameterizations of the IOU process, where the degree of derivative tracking was weak ($\alpha = 20.7$ and $\omega = 1$). The only IOU parameterization where all estimation processes converged was $[\alpha \omega]$. The number of failed estimations due to a non-positive definite information matrix, so that standard errors could not be calculated, was 374,177,13 and 411 for IOU parameterizations $[\alpha \tau]$, $[\ln \alpha \tau]$, $[\alpha^{-2} \tau]$ and $[\alpha^{-2} \omega]$ respectively. When examining these failed estimation processes, the estimate for α was considerably larger than its true value (e.g. estimate was 51.5 when the truth was 20.7) and the likelihood slice for α , with the other parameters held constant at their REML estimates, was flat around the REML estimate of α . For parameterizations $[\ln \alpha \omega]$ and $[\alpha^{-2} \omega]$ seven estimation processes failed because either the NR algorithm did not reach a solution within 100 iterations or the NR algorithm failed due to a non-positive definite Hessian matrix.

Among the estimation processes that converged, the median number of iterations to reach a solution for IOU parameterization $[\alpha^{-2} \tau]$ was more than double the median number for the other IOU parameterizations.

Consider the estimation results of the variance parameters for IOU parameterization $[\alpha \omega]$. The estimates of the random intercept variance, $\text{Var}[u_{0i}]$, were unbiased and the magnitude of the bias of ω was small. Confidence interval coverage for both parameters was close to the nominal level. The distribution of the estimates of α and the distribution of its standard errors (of $\ln \alpha$) were bimodal. In 469 datasets α was overestimated (bias 28.9) and the standard errors of $\ln \alpha$ were very large (i.e. median standard error was 14.5 and the standard deviation of the estimates of $\ln \alpha$ was 0.0972), giving 100% confidence interval coverage. An examination of the likelihood slice for α showed a nearly flat likelihood around the REML estimate of α , which would explain the large standard error. In the remaining

Table 1. Comparing parameterizations of the IOU process for balanced data ($m = 1000$, $n = 20$ and $k = 3$) simulated under the random intercept IOU model with weak derivative tracking ($\alpha = 20.7$, $\omega = 1$).

		Parameterizations					
		$\alpha\tau$	$\alpha\omega$	$\ln \alpha\tau$	$\ln \alpha\omega$	$\alpha^{-2}\tau$	$\alpha^{-2}\omega$
Converged	N	626	1000	823	993	987	582
Opt. Iter.	Median (IQR)	8(3)	9(12)	12(29)	9(7)	25(54)	8(4)
$\text{Var}[u_{0i}]$	Bias $\times 1000$ (MCSE $\times 1000$)	9.3(0.78)	0.45(0.68)	3.9(0.73)	0.18(0.69)	1.3(0.67)	10(0.8)
	Empirical SD	0.083	0.096	0.091	0.097	0.093	0.081
	Median SE (IQR)	0.083(0.07)	0.2(0.66)	0.1(1.3)	0.2(0.74)	0.18(0.94)	0.081(0.053)
	CP %	91%	94%	93%	94%	94%	90%
$\hat{\alpha}$	Bias $\times 1000$ (MCSE $\times 1000$)	-1400(590)	9600(590)	4200(570)	11000(640)	6400(490)	-3700(500)
	Empirical SD	0.55	0.70	0.64	0.73	0.63	0.47
	Median SE (IQR)	0.4(0.85)	1.9(14)	0.71(27)	2(16)	1.5(17)	0.38(0.62)
	CP %	68%	80%	76%	80%	80%	66%
$\hat{\omega}$	Bias $\times 1000$ (MCSE $\times 1000$)	11(0.9)	6.4(0.70)	8.3(0.78)	6.3(0.70)	6.3(0.70)	12(0.94)
	Empirical SD	0.022	0.022	0.022	0.022	0.022	0.022
	Median SE (IQR)	0.025(0.00098)	0.024(0.0059)	0.025(0.006)	0.024(0.0059)	0.024(0.0059)	0.025(0.00084)
	CP %	95%	95%	95%	95%	95%	94%
$\hat{\sigma}$	Bias $\times 1000$ (MCSE $\times 1000$)	17(0.97)	-1.1(0.95)	5.9(1)	-1.8(0.98)	0.98(0.9)	20(0.88)
	Empirical SD	0.1	0.13	0.12	0.13	0.12	0.089
	Median SE (IQR)	0.089(0.18)	0.41(1.7)	0.16(3.5)	0.42(1.9)	0.33(2.3)	0.086(0.14)
	CP %	69%	81%	76%	81%	80%	67%

Optimization performed using NR algorithm. Over 1000 simulated datasets: number converged, median and IQR of the number of optimization iterations (Opt. Iter.) and for each variance parameter v : bias of the estimates of v with Monte Carlo standard error (MCSE), standard deviation of the estimates of its transformation \tilde{v} , median and IQR of the estimated standard error (SE) of \tilde{v} and empirical coverage percentage (CP) of the 95% confidence interval for v . Results reported to two significant figures.

531 datasets α was underestimated (bias -7.52) and the standard errors of $\ln \alpha$ were moderate (i.e. median standard error was 0.352 and the standard deviation of the estimates of $\ln \alpha$ was 0.233), giving 62% confidence interval coverage. So, over the 1000 datasets there was 80% coverage for α . For the same reason the confidence intervals for σ undercovered. For the other five IOU parameterizations the results of the variance parameters followed similar patterns as noted for $[\alpha \omega]$.

When derivative tracking was moderate ($\alpha = 3$, $\omega = 0.5$) all estimation processes converged for IOU parameterizations involving τ and, for IOU parameterizations involving ω between 18 (for $[\alpha \omega]$) and 28 estimation processes did not converge due to failure of the NR algorithm. When derivative tracking was strong ($\alpha = 1.31$, $\omega = 0.1$) all estimation processes converged for all IOU parameterizations. For strong and moderate derivative tracking the medium number of iterations to reach convergence was between 4 and 7 and the results of the variance parameters were very similar across the 6 IOU parameterizations. The estimates of $\text{Var}[u_{0i}]$, ω and σ were unbiased, and under moderate derivative tracking α was unbiased whereas under strong derivative tracking α was only slightly biased (i.e. largest magnitude was 0.0115). For all variance parameters confidence interval coverage was close to the nominal level (95–96 %).

In summary, for weak derivative tracking the preferred IOU parameterization was $[\alpha \omega]$, for moderate derivative tracking the preferred IOU parameterization was either $[\alpha \tau]$, $[\ln \alpha \tau]$ or $[\alpha^{-2} \tau]$, and for strong derivative tracking all IOU parameterizations performed equally well. For the remaining comparisons we estimated the model with respect to the $[\alpha \omega]$ parameterization.

4.1.2. Data structures

Comparison 2(a): Number of subjects versus number of measurements: We estimated the model's parameters when applied to two data structures with 5000 observations in total: $m = 1000$ subjects each with $n = 5$ observations and $m = 250$ subjects each with $n = 20$ observations. Under weak derivative tracking all estimation processes converged for both data structures. The variance parameter results followed the same patterns as described in Section 4.1.1 for IOU parameterization $[\alpha \omega]$ (column 4 of table 1; data structure $[m = 1000 \ n = 20]$). The only notable difference between the data structures was that the confidence interval coverage for $\text{Var}[\hat{u}_{0i}]$ was 89% for $[m = 1000 \ n = 5]$ and 96% for $[m = 250 \ n = 20]$. Under strong derivative tracking all models converged for data structure $[m = 250 \ n = 20]$, all variance estimates were either unbiased or slightly biased and confidence interval coverage was between 94% and 96%. However, for data structure $[m = 1000 \ n = 5]$ 114 estimation processes did not converge within 100 iterations of the NR algorithm. (The same problem occurred, 117 failed estimations, when we used alternative starting values for the optimization algorithm, which were generated assuming strong derivative tracking; that is, the starting values for \mathbf{G} and σ were estimated from a standard linear mixed model, and we set $\alpha = 1$ and $\omega = 0.1$.) Among the converged processes the estimates for α and ω were biased and confidence interval coverage was 100% and 88% respectively. The estimates for $\text{Var}[u_{0i}]$ and σ were unbiased with close to nominal confidence interval coverage.

We also applied the model to smaller datasets (of 500 observations in total): $m = 100$ subjects each with $n = 5$ observations and $m = 250$ subjects each with $n = 20$ observations. Under weak derivative tracking, estimation processes were less likely to converge when the number of subjects was low (4 estimations failed for $[m = 100 \ n = 5]$ compared to 51 failures for $[m = 25 \ n = 20]$). However, under strong derivative tracking estimation processes were less likely to converge when the number of measurements per subject was low (107 estimations failed for $[m = 100 \ n = 5]$ compared to five failures for $[m = 25 \ n = 20]$). The estimation results of the variance parameters followed the same patterns as noted for the data structures with 5000 observations.

In summary, for datasets of 5000 and 500 observations, the variance parameter results were poorer when the number of measurements per subject was 5 compared to 20 measurements per subject.

Comparison 2(b): Time intervals between consecutive measurements: For datasets of 1000 subjects each with 20 observations, we compared estimation of the model when measurements were spaced monthly ($k = 1$), every three months ($k = 3$) and yearly ($k = 12$). The IOU covariance

matrix, \mathbf{H}_i , was the same for all three comparisons and equivalent to weak derivative tracking ($\alpha = 20.7, \omega = 1$ with $k = 1$). All estimation processes converged for $k = 1$ but 10 and 53 estimations failed for $k = 3$ and $k = 12$ respectively. The estimates of $\text{Var}[\hat{u}_{0j}]$ were unbiased for $k = 1, 3$ and 12. All estimates of α were biased and the magnitude of this bias decreased with increasing time interval (i.e. 0.562, 0.199 and 0.0546 for $k = 1, 3$ and $k = 12$ respectively). Note that, the true value of α was 20.7, 6.9 and 1.725 for $k = 1, 3$ and 12 respectively and, as seen in Section 4.1.1, the magnitude of the bias of α was proportional to its true value. Therefore, the magnitude of this bias decreased with increasing time interval because the true value of α decreased with increasing time interval. For all time intervals the estimates of ω were unbiased and the estimates for σ were slightly biased (largest value 0.0045). Confidence interval coverage was close to 95% for all parameters and all time intervals.

For moderately sized datasets of 100 subjects each with five observations, all estimation processes converged for $k = 1$ but 7 and 1 estimations failed for $k = 3$ and $k = 12$ respectively. The parameter results followed the same patterns as described for the larger datasets, but the differences in the magnitude of the biases of α were larger.

In summary, convergence was most successful when measurements were spaced at monthly intervals. Only the results of parameter α differed between the time intervals, where a larger bias occurred for higher values of α .

4.1.3. Optimization algorithms

For datasets of 1000 subjects each with 20 observations spaced at 3 monthly intervals, we compared estimation of the model using different optimization algorithms. Under weak derivative tracking, all estimation processes converged when using only the NR algorithm. There were convergence problems with the combination algorithms: 38 estimations failed when the optimizing algorithm was 3 iterations of FS then NR and 60 estimations failed for 3 iterations of AI then NR. Convergence improved when the number of FS and AI iterations preceding NR were increased from 3 to 10, with only 1 failed estimation for 10 iterations of FS then NR and 13 failed estimations for 10 iterations of AI then NR. The median time to convergence was 54 (IQR 89) seconds for NR only, 118 (IQR 93) seconds for 10 iterations of FS then NR, and 136 (IQR 88) seconds for 10 iterations of AI then NR. The results of the parameter estimates of all combination algorithms were almost identical to the results for estimation using only the NR algorithm (column 4 of Table 1). Under strong derivative tracking all estimation processes converged for all optimization algorithms (NR algorithm only and all combination algorithms). The only notable difference between the optimization algorithms was the median time to convergence: 20 (IQR 9) seconds for NR only, 99 (IQR 8) seconds for 10 iterations of FS then NR, and 47 (IQR 4) seconds for 10 iterations of AI then NR.

For moderately sized datasets ($m = 100, n = 5, k = 3$) with weak derivative tracking the NR only algorithm had fewer failed estimations: 4 estimations failed for NR only, 10 and 9 estimations failed respectively for 3 and 10 iterations of FS then NR, and 29 and 30 estimations failed respectively for 3 and 10 iterations of AI then NR. The median time to convergence was 1 (IQR 1) second for NR only and 3 (IQR 1) seconds for 10 iterations of FS or AI then NR. When the derivative tracking was strong the algorithm with the least number of failed estimations was 10 iterations of AI then NR: 107 failed estimations for NR only, 85 and 73 failed estimations respectively for 3 and 10 iterations of FS then NR, and 76 and 65 failed estimations respectively for 3 and 10 iterations of AI then NR. Among the 107 datasets where the NR only algorithm failed the median starting value of α was 10 (IQR 0), where the true value was 1.31. (The starting values for the other parameters were reasonably close to the truth). When we started the NR algorithm using the alternative starting values (generated assuming very strong derivative tracking) the number of failed estimations for the NR only algorithm decreased to 79. Among the converged estimation processes, the median time to convergence was 2 (IQR 10) seconds for NR only, 4 (IQR 11) seconds for 10 iterations of FS then NR and 3 (IQR 11) seconds for 10 iterations of AI then NR.

In summary, for large datasets with regular and frequent measurements the preferred optimization algorithm was NR only, with respect to convergence and the time taken to reach convergence (on average NR only took at least half the time to reach convergence compared to 10 iterations of FS or AI then NR). In circumstances, such as a moderately sized dataset with a low number of measurements per subject, where starting values are far from the truth the combination of 10 iterations of FS or AI and then NR can improve convergence.

4.1.4. Random slope IOU models

Table 2 presents the results of estimating random slope IOU models when derivative tracking was weak. When the magnitude of the variances and covariance of the random intercept and random slope were small (column RE1 in the table) 122 estimation processes failed to converge: 5 estimations did not converge within 100 NR iterations, 48 NR algorithms failed due to a non-positive definite Hessian matrix, and 69 estimations failed because the information matrix was not positive definite (and so standard errors could not be calculated). For the 122 failed estimations we re-estimated the model using the alternative starting values (generated assuming very strong derivative tracking). Out of these 122 re-estimations 116 converged and the estimated values for the covariance term were very close to zero (median 1×10^{-27}). As described by the Stata manual, when an estimated variance component is close to 0 ‘a ridge is formed by an interval of values near 0, which produce the same likelihood and look equally good to the optimizer’ [10]. When this occurs, then the estimation process can fail to

Table 2. Comparing random slope IOU model for balanced data ($m = 1000, n = 20$ and $k = 3$) simulated with weak derivative tracking ($\alpha = 20.7, \omega = 1$).

		Random slope IOU models		
		RE1	RE2	RE3
Converged	<i>N</i>	878	1000	1000
Opt. Iter.	Median (IQR)	20(14)	12(13)	11(13)
Var[\hat{u}_{0i}]	Bias $\times 1000$ (MCSE $\times 1000$)	0.3(0.72)	0.97(1.1)	0.91(1.8)
	Empirical SD of $\check{\nu}$	0.095	0.036	0.028
	Median SE of $\check{\nu}$ (IQR)	0.23(0.64)	0.055(0.12)	0.035(0.05)
	CP %	95%	97%	97%
Cov[$\hat{u}_{0i}, \hat{u}_{1i}$]	Bias $\times 1000$ (MCSE $\times 1000$)	-0.26(0.31)	0.28(0.65)	0.44(1.1)
	Empirical SD	2.1	0.076	0.053
	Median SE (IQR)	0.51(5.1)	0.093(0.17)	0.059(0.068)
	CP %	95%	96%	96%
Var[\hat{u}_{1i}]	Bias $\times 1000$ (MCSE $\times 1000$)	0.59(0.3)	-2.4(0.68)	-2.8(1)
	Empirical SD	0.97	0.043	0.032
	Median SE (IQR)	0.57(0.831)	0.043(0.0031)	0.032(0.0011)
	CP %	93%	95%	95%
$\hat{\alpha}$	Bias $\times 1000$ (MCSE $\times 1000$)	9100(610)	8600(580)	8800(580)
	Empirical SD	0.68	0.70	0.70
	Median SE (IQR)	2.5(14)	1.7(14)	1.7(14)
	CP %	81%	79%	79%
$\hat{\omega}$	Bias $\times 1000$ (MCSE $\times 1000$)	5.9(0.79)	9(0.81)	9.1(0.82)
	Empirical SD	0.023	0.025	0.025
	Median SE (IQR)	0.025(0.008)	0.029(0.0088)	0.029(0.0089)
	CP %	96%	96%	95%
$\hat{\sigma}$	Bias $\times 1000$ (MCSE $\times 1000$)	-1.3(0.98)	0.16(0.95)	0.01(0.96)
	Empirical SD	0.12	0.13	0.13
	Median SE (IQR)	0.48(1.7)	0.37(1.7)	0.36(1.7)
	CP %	82%	79%	79%

Note: Optimization performed using NR algorithm and IOU parameterization [$\alpha \omega$]. Over 1000 simulated datasets: number converged, median and IQR of the number of optimization iterations (Opt. Iter.) and for each variance parameter ν : bias of the estimates of ν with Monte Carlo standard error (MCSE), standard deviation of the estimates of its transformation $\check{\nu}$, median and IQR of the estimated standard error (SE) of $\check{\nu}$ and empirical coverage percentage (CP) of the 95% confidence interval for ν . Results reported to two significant figures.

converge within a reasonable number of iterations, or fail due to an unstable Hessian matrix, or fail to calculate reliable standard errors [10]. All estimation processes converged when the random effects' variances and covariances were larger (columns RE2 and RE3) and the median number of iterations to reach convergence was almost half the median number required when the random effects were small. For all random slope IOU models, the random effects (co)variance estimates and estimates for ω were mostly unbiased (slightly biased in one case) and confidence interval coverage was close to nominal. The results for α and σ follow the same patterns as noted for the random intercept model when the true value of α was large (see Section 4.1.1).

When derivative tracking was moderate and strong convergence problems were more common for the random slope IOU model with small random effects variances and covariance (RE1) than for the models with larger random effects' variances and covariance (RE2 and RE3): for moderate derivative tracking 84,25 and 16 estimations failed for RE1, RE2 and RE3 respectively and for strong derivative tracking 10 estimations failed for RE1 and all estimations converged for RE2 and RE3. (Under moderate derivative tracking, convergence improved for all models when we estimated with respect to IOU parameterization $[\alpha \ \tau]$. There were more failed estimations for the RE1 model; models RE2 and RE3 had one failed estimation). With respect to the variance parameter results, under moderate derivative tracking, the only notable difference between the random slope IOU models was that the confidence interval coverage for the variance of the random slope term, $\text{Var}[u_{1i}]$, was 89% for RE1, and 95% for RE2 and RE3. This confidence interval undercoverage was due to underestimation of the standard errors. Otherwise, under moderate and strong derivative tracking, all parameter estimates were either unbiased or slightly biased (magnitude less than 0.017) and confidence intervals had close to nominal coverage.

In summary, the linear IOU mixed model with a random slope effect can be successfully estimated provided the random effects variances and covariances are not close to zero.

4.2. Unbalanced longitudinal data

For unbalanced data generated under the *intermittent dropout* scenario, on average (i.e. taking means over 1000 datasets), the median (IQR), minimum and maximum follow-up time were respectively 4.5 years (IQR 2.4 years), 0.86 and 6.3 years, and the median (IQR), minimum and maximum time interval between consecutive measurements were respectively 3 months (IQR 1.4 months), 1 week and 2.6 years. The number of failed estimations was 78, 28 and 207 for weak, moderate and strong derivative tracking respectively. The estimation processes failed because either the NR algorithm did not reach a solution within 100 iterations or the information matrix was not positive definite. To the same simulated datasets we re-estimated the model using the alternative starting values. For data simulated under strong derivative tracking all 1000 estimation processes then converged. This improvement in convergence occurred because the alternative starting values were closer to the truth than our original starting values. However, for weak and moderate derivative tracking, our original starting values were closer to the truth and so there was no improvement in convergence using these alternative starting values.

Under the *target visit* scenario, on average, the median (IQR), minimum and maximum follow-up time were respectively 3.12 years (IQR 3.29 years), 3.20 months and 6.07 years, and the median (IQR), minimum and maximum time interval between consecutive measurements were respectively 3.00 months (IQR 1.34 months), 1.00 week and 7.01 months. Given the same degree of derivative tracking, the number of failed estimation processes was lower for the *intermittent dropout* scenario than for the *target visits* scenario (106, 118 and 417 failed estimations for the *target visits* scenario under weak, moderate and strong derivative tracking respectively). Successful convergence of the linear mixed model depends upon the number of observations and coverage of the continuous time line of measurements (i.e. from natural time zero to end of follow-up), where an estimation process is more likely to converge the greater the coverage level. Compared to the *target visits* scenario, there was greater variability (between and within subjects) in the times of the measurements for the *intermittent*

dropout scenario. Therefore, since data simulated under both scenarios had the same number of measurements, the level of coverage of the continuous time line was higher for the *intermittent dropout* scenario, and hence a greater number of converged estimations.

Among the converged estimations, the median number of iterations to reach convergence (5–8) and, the estimation results of the fixed effects and variance parameters were very similar between the *target visits* and *intermittent dropout* scenarios. The fixed effects estimates were unbiased and confidence interval coverage was between 93% and 96% for weak, moderate and strong derivative tracking. Under weak derivative tracking, the estimates of $\text{Var}[\hat{u}_{0i}]$ were unbiased and only slightly biased for ω , and confidence interval coverage was 96%. The magnitude of the bias for α was at most 1.06 and confidence interval coverage for α and σ was 92–93%. For moderate and strong derivative tracking, all variance estimates were either unbiased or slightly biased (e.g. 0.00313 for α) and confidence interval coverage was close to the nominal level.

5. Application

We applied the linear mixed IOU model to a large dataset of longitudinal height measurements in 856 boys, aged 9–18 years, who attended Christ's Hospital School between 1936 and 1969 [35]. The median number of measurements per boy was 60 (IQR 16) and the median time interval between consecutive measurements (within a boy) was 1.3 months (IQR 0.26).

Figure 1 shows the observed growth trajectories for 10 boys. As the shape of the growth trajectories were close to linear we decided to fit a linear mixed IOU model with a linear population slope.

Lampel et al. [36] state that growth in height in individuals does not necessarily follow a simple trajectory but instead proceeds in a series of jumps separated by period of zero growth. The linear mixed IOU model is ideally suited to allow for variable gradients without the need to specify a complicated function for the growth trajectory. Also, given the frequency of measurements, the linear mixed IOU model allows for any residual serial correlation.

We compared the model fit of a standard linear mixed with a random intercept and random slope, a random intercept IOU model and a random slope IOU model. All models had the same fixed effects and the models with a random slope had an unstructured random effects covariance matrix. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for the linear mixed model, random intercept IOU model and random slope IOU model were respectively $\text{AIC} = -157,005$ $\text{BIC} = -156,954$, $\text{AIC} = -216,254$ $\text{BIC} = -216,203$ and $\text{AIC} = -216,397$

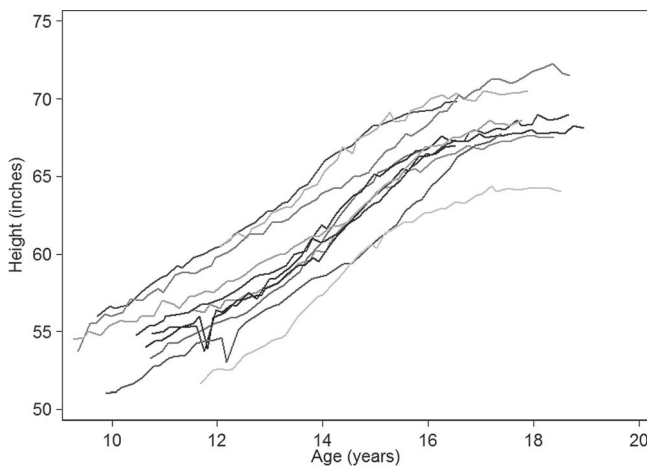


Figure 1. Line plots of height measurements from 10 students who attended Christ's Hospital School between 1936 and 1969.

Table 3. Fixed effects and variance parameter results of three models fitted to longitudinal height measurements (in metres) from Christ’s Hospital School: the standard linear mixed model with a random intercept and slope (RS RE model), the random intercept IOU model (RI IOU model) and the random slope IOU model (RS IOU model). Results reported to two significant figures.

	RS RE model		RI IOU model		RS IOU model	
	Estimate	Confidence Interval	Estimate	Confidence Interval	Estimate	Confidence Interval
Age (years)	0.055	[0.054, 0.055]	0.046	[0.045, 0.046]	0.046	[0.045, 0.047]
Constant	0.83	[0.82, 0.84]	0.94	[0.93, 0.95]	0.93	[0.92, 0.95]
$\text{Var}[\hat{u}_{0i}]$	0.019	[0.018, 0.021]	0.0073	[0.0063, 0.0084]	0.011	[0.0093, 0.013]
$\text{Cov}[\hat{u}_{0i}, \hat{u}_{1i}]$	-0.00087	[-0.00098, -0.00077]	–	–	-0.00054	[-0.00066, -0.00041]
$\text{Var}[\hat{u}_{1i}]$	0.000063	[0.000057, 0.000070]	–	–	0.000026	[0.000019, 0.000037]
$\hat{\alpha}$	–	–	1	[0.95, 1.1]	1	[0.98, 1.1]
$\hat{\omega}$	–	–	0.0015	[0.0014, 0.0016]	0.0015	[0.0014, 0.0016]
$\hat{\sigma}$	0.026	[0.025, 0.026]	0.0086	[0.0086, 0.0087]	0.0086	[0.0086, 0.0087]

BIC = -216,329. Based on the AIC and BIC values the model with the best fit to the data was the random slope IOU model.

Table 3 presents the results of all three models. The time scale was age in years and the natural time zero was birth. Therefore, the constant fixed effect denotes the average height (in metres) of the boys at birth. The estimates $\hat{\alpha}$ and $\hat{\omega}$ indicate strong derivative tracking. Comparing the results of the standard linear mixed model (RS RE model) and the random slope IOU model we note that the fixed effects estimates differ and the corresponding confidence intervals do not overlap. Given the small magnitude of the estimates for the random effects variances and covariances there was little difference in the fixed effects results and the remaining parameter results between the random intercept IOU model (RI IOU model) and the random slope IOU model.

6. Discussion

We have conducted the first evaluation of the feasibility and practicality of REML estimation of the linear mixed IOU model, for both balanced and unbalanced data. Our simulation study demonstrated that the preferred parameterization of the IOU process was $[\alpha \omega]$ even when the true value of α was large. If convergence problems occur with parameterization $[\alpha \omega]$ then the user could try parameterization $[\alpha \tau]$, particularly when derivative tracking is neither weak or strong. Estimation of the linear mixed IOU model is feasible for large datasets (i.e. 20,000 observations in total) and moderately sized datasets (i.e. 500 observations). For datasets of 500 observations the model was successfully fitted to data where each subject had 5 measurements each and also to data where the number of subjects was 25. The NR optimization algorithm was preferred over FS and AI algorithms for datasets with regular and frequent measurements. In circumstances, such as a moderately sized dataset with a low number of measurements per subject, where starting values are far from the truth, the combination of 10 iterations of FS or AI and then NR can improve convergence. When applied to a large dataset, the convergence and estimation properties of the random slope IOU model were similar to the random intercept IOU model provided the magnitude of the variance for the random slope effect and its covariance with the random intercept were larger than 0.01. We have demonstrated that it is feasible to apply the model to large unbalanced data with (non-informative) and intermittent dropout.

In our example the linear mixed IOU model was a better fit to the data than the standard linear mixed model and, importantly, inference for the fixed effects coefficients differed between the two models. Taylor et al. [5] showed that the linear mixed IOU model and the Brownian Motion mixed model gave a much better fit to the data (from the Multicenter AIDS Cohort Study) than the standard linear mixed model, but concluded that the choice of the covariance structure made little difference to the point estimates and standard errors of the fixed effects. Using simulations, Taylor and Law [9] compared the linear mixed IOU model, the Brownian Motion mixed model and the standard linear mixed model with respect to the robustness of subject-level predictions when the covariance structure

was incorrectly specified. They concluded that the linear mixed IOU model was the preferred model because of its efficiency and robustness properties.

Our method for deriving starting values can only be used when the variance structure is that of the random intercept IOU model or the random slope IOU model. We also considered a general method that can be applied to all linear mixed IOU models, which assumes strong derivative tracking. The rationale behind this method is analogous to the way some commercial software (e.g. SAS, Stata) chooses starting values for a conditional independence model (for example the starting value for σ^2 is obtained by fitting a linear regression model to the data using ordinary least squares (OLS) and starting values for the parameters of \mathbf{G} are either 0 or a multiple of the OLS estimate for σ^2). For the linear mixed IOU model we first fitted a conditional independence model (using the EM algorithm). The estimates from this model then form the starting values for the measurement error variance $\hat{\sigma}^2$ and the random effect parameters of \mathbf{G} . The starting values of α and ω were set to 1 and 0.1 respectively.

Estimation problems occurred when derivative tracking was weak with large α . The most likely cause was a flat likelihood surface near the optimum so that either the optimization algorithm had difficulties converging to a global optimum or it converged with an observed information matrix that was not positive definite, and so standard errors (for the variance parameters) could not be computed. Convergence with valid standard errors given a flat likelihood surface was most likely when the dataset was large and the IOU process was parameterized as $[\alpha \ \omega]$. However, estimates of α were substantially overestimated with large standard errors (since they were computed from the reciprocal of the curvature of the REML log-likelihood at the optimum). Taylor et al. [5] also reported a very flat likelihood surface for estimates of $\alpha > 2.5$ with profile likelihood 95% confidence intervals where the upper limit was ∞ , with $\alpha = \infty$ corresponding to one of the special limiting cases of the linear mixed IOU model: the Brownian Motion mixed-effects model. For an estimation process that converges one can diagnose a flat likelihood surface by starting the estimation process with two or more different sets of starting values. If the resulting sets of estimates differ greatly then this indicates a flat likelihood surface near the optimum [37]. The strength of the derivative tracking, as mentioned earlier, depends upon the values of α and ω and the time intervals between consecutive measurements. For large datasets, these numerical problems did not occur for the scenario of monthly measurements with $\alpha = 20.7$ and $\omega = 1$; however, these problems were prominent when measurements were every 3 months and derivative tracking was $\alpha = 20.7$ and $\omega = 1$, which is equivalent to an IOU covariance matrix based on monthly measurements with $\alpha = 62.1$ and $\omega = 3$.

Convergence problems may be overcome by simplifying the model; for example, removing one or more random effects (apart from the random intercept) from the model, or fitting the Brownian Motion mixed model, which has one less parameter than the linear mixed IOU model, and may be preferable when derivative tracking is suspected to be weak. When convergence problems are suspected to be caused by difficulties in estimating α , one suggestion is to derive starting values (of the variance parameters) that are closer to the truth using a computer-intensive approach as follows: first, select values for α over an interval that is expected to contain its true value. Second, for each α value: fix α equal to this value and perform REML estimation with respect to the remaining variance parameters, and record the deviance ($-2 \times$ REML log-likelihood). Third, for the model with the lowest deviance use the α value and the corresponding estimated parameters as starting values and refit the model where all variance parameters, including α , are estimated. This is a simple and, almost, automatic approach that can be easily implemented. However, the approach is likely to be time consuming, especially when searching over many potential values for α . Also, setting the interval of α to be too narrow and/or too coarse (i.e. large gaps between consecutive α values) may omit the true (or close to true) α value, thereby generating poor starting values.

Future work could include: (1) a comparison of the fit and subject-level predictions of the linear mixed IOU model and the Brownian Motion mixed model when derivative tracking is very weak and the likelihood surface of the linear mixed IOU model tends to be flat; (2) Evaluation of the estimation and fit of the linear mixed IOU model when the population slope is nonlinear (e.g. slope described

by a fractional polynomial); and (3) provision of software for the estimation of the Brownian Motion mixed model.

The linear mixed IOU model is a flexible modelling approach but makes parametric assumptions that may not always be well supported by the data. Alternative, non-parametric or semiparametric methods, may be preferred in some situations. For example, the Functional Data Analysis (FDA) approach to longitudinal data provides non-parametric methods, such as smoothing splines, for modelling trajectories [38], where the trajectory (or curve) is the basic unit of data analysis. In general, FDA requires a large number of regularly spaced measurements, such as data on electrical activity of the heart (electrocardiography), electrical activity along the scalp (electroencephalography), and continuous activity monitoring through accelerometers. [39]. However, the development of FDA methodology for longitudinal data, with relatively small numbers of measurements per subject and irregularly spaced, is an area of ongoing research [40], and has been used to model trajectories of prostate specific antigen (a biomarker of prostate cancer) [41, 42]. Also, under certain conditions, the linear mixed model corresponds to a smoothing spline model [43–45], and this connection has been used to develop methods that exploit both features of linear mixed models and FDA [43, 46].

In conclusion, the linear mixed IOU model can provide a better fit to the data than the standard linear mixed model with estimates and confidence intervals for the fixed effects. We recommend estimation of the linear mixed IOU model with respect to the $[\alpha \omega]$ parameterization and using the NR algorithm. If estimation problems occur the user can try our alternative method for deriving starting values and/or estimation with respect to IOU parameterization $[\alpha \tau]$. The model can be applied to moderately sized balanced datasets and large balanced and unbalanced datasets. However, estimation problems may occur when the number of measurements per subject is less than 5. One can successfully estimate a linear IOU mixed model with a random slope effect, provided the random effects variances and covariances are not close to zero. When derivative tracking is weak it is important to check for a flat likelihood surface. We have created a Stata command `xtmixediou`, that fits the linear mixed IOU model, and an accompanying command, `predict`, to generate predictions under the model. Both of these commands are publicly available from the Statistical Software Components archive.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Rachael Hughes was supported by Medical Research Council grant [MR/J013773/1] and Jonathan Sterne was supported by National Institute for Health Research Senior Investigator award [NF-SI-0611-10168]. We acknowledge Professor Yoav Ben-Shlomo from the School of Social and Community Medicine, University of Bristol for granting access to data from Christ's Hospital School study.

References

- [1] Laird N, Ware J. Random-effects models for longitudinal data. *Biometrics*. 1982;38:963–974.
- [2] Boscardin W, Taylor J, Law N. Longitudinal models for AIDS marker data. *Stat Methods Med Res*. 1998;7(1):13–27.
- [3] Sowers ME, Crutchfield M, Randolph JF, et al. Urinary ovarian and gonadotrophin hormone levels in premenopausal women with low bone mass. *J Bone Miner Res*. 1998;13(7):1191–1202.
- [4] Diggle P. An approach to the analysis of repeated measurements. *Biometrics*. 1988;44(4):959–971.
- [5] Taylor J, Cumberland W, Sy J. A stochastic model for analysis of longitudinal AIDS data. *J Am Stat Assoc*. 1994;89(427):727–736.
- [6] Zhang D, Lin X, Raz J, et al. Semiparametric stochastic mixed models for longitudinal data. *J Am Stat Assoc*. 1998;93:710–719.
- [7] Stirrup O, Babiker A, Carpenter J, et al. Fractional Brownian motion and multivariate-t models for longitudinal biomedical data, with application to CD4 counts in HIV-positive patients. *Stat Med*. 2016;35(9):1514–1532.

- [8] Wolbers M, Babiker A, Sabin C. Pretreatment CD4 cell slope and progression to AIDS or death in HIV-infected patients initiating antiretroviral therapy – The CASCADE Collaboration: a collaboration of 23 cohort studies. *PLOS Med.* 2010;7(2):1–9.
- [9] Taylor J, Law N. Does the covariance structure matter in longitudinal modelling for the prediction of future CD4 counts?. *Stat Med.* 1998;17:2381–2394.
- [10] StataCorp. *Stata statistical software: release 13.* College Station (TX): StataCorp LP; 2013.
- [11] Patterson H, Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika.* 1971;58(3):545–554.
- [12] Pinheiro J, Bates D. *Mixed effects models in S and Splus.* New York: Springer-Verlag; 2000.
- [13] Diggle P, Heagerty P, Liang KY, et al. *Analysis of longitudinal data.* 2nd ed. Oxford: Oxford University Press; 2002.
- [14] Wu L. *Mixed effects models for complex data.* Boca Raton: Chapman and Hall/CRC; 2009.
- [15] Rubin D. Inference and missing data (with discussion). *Biometrika.* 1976;63(3):581–592.
- [16] Lindstrom M, Bates D. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures. *J Am Stat Assoc.* 1988;83:1014–1022.
- [17] Jennrich R, Sampson P. Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics.* 1976;18(1):11–17.
- [18] Gilmour A, Thompson R, Cullis B. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics.* 1995;51(4):1440–1450.
- [19] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B.* 1977;39:1–38.
- [20] Gumedze F, Dunne T. Parameter estimation and inference in the linear mixed model. *Linear Algebra Appl.* 2011;435(8):1920–1944.
- [21] Meng XL, Rubin D. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika.* 1993;80(2):267–278.
- [22] Meng XL, Dykvan D. Fast EM-type implementations for mixed effects models. *J R Stat Soc B.* 1998;60(3):559–578.
- [23] Dykvan D. Fitting mixed-effects models using efficient EM-type algorithms. *J Comput Graph Stat.* 2000;9:78–98.
- [24] Liu C, Rubin D, Wu Y. Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika.* 1998;85(4):755–770.
- [25] Foulley J, Dykvan D. The PX-EM algorithm for fast stable fitting of Henderson’s mixed model. *Genet Sel Evol.* 2000;32(2):143–163.
- [26] Pinheiro J, Bates D. Unconstrained parametrizations for variance-covariance matrices. *Stat Comput.* 1996;6(3):289–296.
- [27] Wolfinger R, Tobias R, Sall J. Computing Gaussian likelihoods and their derivatives for general linear mixed models. *SIAM J Sci Comput.* 1994;15(6):1294–1310.
- [28] Goodnight J, Hemmerle W. A simplified algorithm for the w transformation in variance component estimation. *Technometrics.* 1979;21(2):265–267.
- [29] Hemmerle W, Hartley H. Computing maximum likelihood estimates for the mixed A. O. V. model using the WTransformation. *Technometrics.* 1973;15:819–831.
- [30] Goodnight J. A tutorial on the SWEEP operator. *Am Stat.* 1979;33:149–158.
- [31] Stata. *Mata reference manual release 13.* StataCorp, College Station, TX: Stata Press; 2013.
- [32] UK Collaborative HIV Cohort Steering Committee. The creation of a large UK-based multicentre cohort of HIV-infected individuals: the UK collaborative HIV cohort (UK CHIC) study. *HIV Med.* 2004;5(2):115–124.
- [33] White I. Simsum: analyses of simulation studies including Monte Carlo error. *Stata J.* 2010;10:369–385.
- [34] Kackar R, Harville D. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J Am Stat Assoc.* 1984;79:853–862.
- [35] Sandhu J, Ben-Shlomo Y, Cole T, et al. The impact of childhood body mass index on timing of puberty, adult stature and obesity: a follow-up study based on adolescent anthropometry recorded at Christ’s Hospital (1936 to 1964). *Int J Obesity.* 2006;30(1):14–22.
- [36] Lampl M, Ashizawa K, Kawabata M, et al. An example of variation and pattern in saltation and stasis growth dynamics. *Ann Hum Biol.* 1998;25(3):203–219.
- [37] Pawitan Y. *In all likelihood.* Oxford: Oxford University Press; 2013.
- [38] Müller HG. *Longitudinal data analysis.* Boca Raton: Chapman and Hall/CRC; 2008.
- [39] Sorensen H, Goldsmith J, Sangallic L. An introduction with medical applications to functional data analysis. *Stat Med.* 2013;32(30):5222–5240.
- [40] Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc.* 2005;100(470):577–590.
- [41] Zhu B, Taylor J, Song PXX. Semiparametric stochastic modeling of the rate function in longitudinal studies. *J Am Stat Assoc.* 2011;106(496):1485–1495.
- [42] Zhu B, Taylor J, Song PXX. Stochastic functional data analysis: a diffusion model-based approach. *Biometrics.* 2011;67(4):1295–1304.

- [43] Brumback B, Rice J. Smoothing spline models for the analysis of nested and crossed samples of curves. *J Am Stat Assoc.* [1998](#);93(443):961–976.
- [44] Rice J. Functional and longitudinal data analysis: perspectives on smoothing. *Stat Sin.* [2004](#);14:631–647.
- [45] Nummi T, Koskela L. Analysis of growth curve data by using cubic smoothing splines. *J Appl Stat.* [2008](#);35(6):681–691.
- [46] Guo W. Functional data analysis in longitudinal settings using smoothing splines. *Stat Methods Med Res.* [2004](#);13:49–62.