HHS Public Access

Author manuscript

J Comput Graph Stat. Author manuscript; available in PMC 2015 March 29.

Published in final edited form as:

J Comput Graph Stat. 2009 ; 18(4): 797–817. doi:10.1198/jcgs.2009.07130.

Fast Implementation for Normal Mixed Effects Models With Censored Response

Florin Vaida [Associate Professor] and **Lin Liu [Assistant Project Scientist]**

Department of Family and Preventive Medicine, UC San Diego School of Medicine, La Jolla, CA 92093-0717

Lin Liu: linliu@ucsd.edu

Abstract

We propose an EM algorithm for computing the maximum likelihood and restricted maximum likelihood for linear and nonlinear mixed effects models with censored response. In contrast with previous developments, this algorithm uses closed-form expressions at the E-step, as opposed to Monte Carlo simulation. These expressions rely on formulas for the mean and variance of a truncated multinormal distribution, and can be computed using available software. This leads to an improvement in the speed of computation of up to an order of magnitude. A wide class of mixed effects models is considered, including the Laird–Ware model, and extensions to different structures for the variance components, heteroscedastic and autocorrelated errors, and multilevel models. We apply the methodology to two case studies from our own biostatistical practice, involving the analysis of longitudinal HIV viral load in two recent AIDS studies.

The proposed algorithm is implemented in the R package `lmec`. An appendix which includes further mathematical details, the R code, and datasets for examples and simulations are available as the online supplements.

Keywords

Detection limit; EM algorithm; HIV viral load; Maximum likelihood; Truncated multinormal distribution

1. INTRODUCTION

Linear and nonlinear mixed effects models (N/LME) are now well established in statistical methodology and practice; see, for example, Davidian and Giltinan (1995); Pinheiro and Bates (2000); Jiang (2007). Statistical software implementations, including the `nlme` and

Supplementary materials for this article are available at http://pubs.amstat.org/toc/jcgs/18/4.

SUPPLEMENTAL MATERIALS

Data Sets, Computer Code, and Appendix: This archive contains datasets from a study of Unstructured Treatment Interruption in HIV-infected adolescents in four institutions in the United States and from a large multicenter observational study of subjects with acute and early HIV infection (AIEDRP), R codes to the models in Sections 5.1 and 5.2 and fitting a nonlinear mixed effects model for censored normal responses. Also there is an appendix to the main article which gives mathematical details of the proposed method. (Supplement.zip, zip archive)

`lme4` suites for R/S-PLUS (Pinheiro et al. 2006; Bates and Sarkar 2007) or PROC NLMIXED in SAS (Wolfinger 1999), are fast and efficient, allowing repeated runs of the procedure in real time, such as for model comparison, multiple model fits, or statistical simulations. Modeling censored observations using N/LME occurs often in biomedical applications, such as pharmacokinetics (PK). In our own experience we encountered the need for these models in analyzing HIV viral load data, where observations occur below (left-censored) or above (right-censored) the limit of quantitation of the assay; see Saitoh et al. (2008) and the case studies in Section 5. Hughes (1999) proposed a Monte Carlo EM algorithm (MCEM) for LME with censored response (LMEC). Vaida, Fitzgerald, and DeGruttola (2007) proposed a hybrid EM (HEM) using a more efficient implementation of Hughes' algorithm, including numeric computation at the E-step for clusters with one or two censored observations. They also extended the algorithm to NLME with censored data (NLMEC). Their MCEM improves the simulation at the E-step, the numeric implementation at the M-step, and includes automatic monitoring and stopping of the algorithm. However, by its nature MCEM is an expensive proposition, due to a combination of Monte Carlo simulation with the iterative procedure (Ruppert 2005). Whereas HEM takes 20–100 sec to compute with satisfactory precision, this is still too much for routine use, as in simulations, or as part of more complex statistical procedures. In this article we propose a numeric implementation of the EM algorithm for N/LMEC with greatly improved speed and precision. We show that the E-step reduces to computing the first two moments of certain truncated multivariate normal (multinormal) distributions. The general formulas for these moments were derived by Tallis (1961) and Finney (1962). They require the multinormal CDF, for which we use the `mvtnorm` package in R (Genz 1992). This implementation computes the maximum likelihood (MLE) or restricted maximum likelihood (REML) estimators. The likelihood function is easily computed as a by-product of the E-step and is used for monitoring convergence and for model selection (AIC, likelihood ratio test). In contrast with the existing literature, we give here explicit derivations for a wide class of mixed effects models with censored response, including the Laird–Ware model and extensions to different structures for the variance components, heteroscedastic and autocorrelated errors, and multilevel models. The method is implemented in the R package `lmec` (Vaida and Liu 2009) available on CRAN.

Section 2 presents the main method. In Sections 3 and 4 the extension to more general LMEC and to NLMEC is discussed. Section 5 includes two case studies of modeling HIV viral load. In Section 6 the new algorithm is compared with the MCEM version in terms of speed and precision, via simulations. A discussion in Section 7 ends the exposition. To simplify the presentation, further mathematical details were included in an online appendix accessible via the journal's webpage, which also provides the R code and datasets for examples and simulations.

2. LINEAR MIXED EFFECTS WITH CENSORED RESPONSE

2.1 Model Specification and Maximum Likelihood Estimation

Consider the general mixed effects model

$$y=X\beta+Zb+e, b\sim N(0,\sigma^2 D), e\sim N(0,\sigma^2\Lambda), \quad (2.1)$$

where β is the vector of fixed effects, b is the vector of random effects, e is the error vector, X ($n \times p$) and Z ($n \times r$) are design matrices, and D and Λ are scaled variance matrices; b and e are assumed independent. The response y is not fully observed for all individual components. Assuming left-censoring, let the observed data be (Q, C), where Q is the n-vector of uncensored readings or censoring levels, as the case may be, and C is the vector of censoring indicators.

We focus first on the Laird–Ware model

$$y_i=X_i\beta+Z_i b_i+e_i, b_i\sim N(0,\sigma^2 F), e_{ij}\sim N(0,\sigma^2), i=1,\ldots,m, \quad (2.2)$$

where X_i ($n_i \times p$) and Z_i ($n_i \times q$) are the design matrices, and b_i and error vectors $e_i = (e_{i1}, \ldots, e_{in_i})^\top$ are independent for all i and independent of each other. Following Magnus and Neudecker (1999) and Wand (2002), let vec(·) denote the function which stacks vectors, or matrices of same number of columns, and diag($A_1, \ldots, A_m$) be the block-diagonal matrix with diagonal blocks $A_1, \ldots, A_m$; if $A_1 = \cdots = A_m = A$, we write $\mathrm{diag}_m(A)$ for diag($A_1, \ldots, A_m$). Then clearly (2.2) is a special case of (2.1), with $y=\mathrm{vec}(y_1,\ldots,y_m)=(y_1^\top,\ldots,y_m^\top)^\top$, X = vec($X_1, \ldots, X_m$), Z = diag($Z_1, \ldots, Z_m$), $r = mq$, b = vec($b_1, \ldots, b_m$), e = vec($e_1, \ldots, e_m$), D = $\mathrm{diag}_m(F)$, $\Lambda = I_n$, and $n=\sum_{i=1}^m n_i$. (Conversely, (2.1) is a special case of (2.2), with m = 1.) F is a positive-definite matrix depending on a vector of parameters γ. Put $\sigma^2 F = \Psi$ and $V_i=\mathrm{var}(y_i)=Z_i\Psi Z_i^\top+\sigma^2 I$.

Write Q = vec($Q_1, \ldots, Q_m$), C = vec($C_1, \ldots, C_m$), such that the observed data for the ith subject is (Q_i, C_i). For individual observations within cluster i we have

$$y_{ij} \le Q_{ij} \text{ if } C_{ij}=1; y_{ij}=Q_{ij} \text{ if } C_{ij}=0. \quad (2.3)$$

The EM algorithm for the Laird–Ware model with censored data was proposed by Hughes (1999), with computational improvements given by Vaida, Fitzgerald, and De-Gruttola (2007). Following the notation of the latter article, let δ = vec(β, b); decompose $F^{-1} = \Delta^\top\Delta$ and define $\tilde{y}$ = vec($\tilde{y}_1, \ldots, \tilde{y}_m$) and

$$\begin{pmatrix} \tilde{y}_i & \tilde{X}_i & \tilde{Z}_i \end{pmatrix} = \begin{pmatrix} y_i & X_i & Z_i \\ 0 & 0 & \Delta \end{pmatrix} \text{ and}$$

$$M = \begin{pmatrix} \tilde{X}_1 & \tilde{Z}_1 & & 0 \\ \vdots & & \ddots & \\ \tilde{X}_m & 0 & & \tilde{Z}_m \end{pmatrix} = \begin{pmatrix} \tilde{X} & \tilde{Z} \end{pmatrix}. \quad (2.4)$$

Let $y^c = \{y_{ij} : C_{ij} = 1\}$ be the set of left-censored observations. In the EM we update β, σ^2 with y^c as missing data, and Ψ using y^c and b as missing data (Vaida, Fitzgerald, and DeGruttola 2007). The M-step updates are:

$$\hat{\delta}=(M^\top M)^{-1}M^\top E(\tilde{y}|Q,C), \quad (2.5)$$

$$\hat{\sigma}^2=\frac{1}{n}\|E(\tilde{y}|Q,C)-M\hat{\delta}\|^2+\frac{1}{n}\sum_{i=1}^{m}\mathrm{tr}\{\mathrm{var}(y_i|Q,C)\}-\frac{1}{n}\sum_{i=1}^{m}\mathrm{tr}\{W_iZ_i^\top\mathrm{var}(y_i|Q,C)Z_i\}, \quad (2.6)$$

where $W_i=(\tilde{Z}_i^\top\tilde{Z}_i)^{-1}$ and $E(y_i \mid Q, C)$, $\mathrm{var}(y_i \mid Q, C)$ are the mean and variance of y_i conditional on Q, C, taken at the current parameter value $\theta = (\beta, \sigma^2, F)$.

The update for *unstructured* Ψ is given by

$$\hat{\Psi}=\frac{1}{m}\sum_{i=1}^{m}E(b_i|Q,C)E(b_i|Q,C)^\top+\frac{1}{m}\sum_{i=1}^{m}\mathrm{var}(b_i|Q,C). \quad (2.7)$$

It can be shown that $E(b_i|Q,C)=W_iZ_i^\top\{E(y_i|Q,C)-X_i\beta\}$, $\mathrm{var}(b_i|Q,C)=\sigma^2W_i+W_iZ_i^\top\mathrm{var}(y_i|Q,C)Z_iW_i$. See Appendix A1 for details. More general structures for Ψ are considered in Section 3. The computations use dimension reduction based on QR decomposition which takes advantage of the sparse nature of the matrix M (Pinheiro and Bates 2000). The key feature is that the number of columns of the matrices to be decomposed does not increase with the number of clusters m or the number of data points n.

From (2.5)–(2.7) it is clear that the E-step reduces to the computation of $E(y_i \mid Q_i, C_i, \theta)$ and $\mathrm{var}(y_i \mid C_i, Q_i, \theta)$, that is, the mean and variance of a truncated multinormal distribution. These can be determined in closed form, as a function of multinormal probabilities, using a sequence of simple transformations as follows.

i. The first step is to treat separately the observed and censored components of y. Partition y_i into the observed and censored parts: $y_i=\mathrm{vec}(y_i^o, y_i^c)$, that is, $C_{ij} = 0$ for all elements in y_i^o, and 1 for all elements in y_i^c; write accordingly $Q_i=\mathrm{vec}(Q_i^o, Q_i^c)$. Ignoring censoring for the moment, we have that marginally $y_i\sim N(X_i\beta, \Sigma=\sigma^2(I+Z_iFZ_i^\top))$. Then $y_i^o\sim N(X_i^o\beta, \Sigma_{oo})$, $y_i^c|y_i^o\sim N(\mu_i, S_i)$, where

$$\mu_i=X_i^c\beta+\Sigma_{co}\Sigma_{oo}^{-1}(y_i^o-X_i^o\beta), \quad S_i=\Sigma_{cc}-\Sigma_{co}\Sigma_{oo}^{-1}\Sigma_{oc},$$

and $\Sigma=\begin{pmatrix}\Sigma_{oo} & \Sigma_{oc}\\ \Sigma_{co} & \Sigma_{cc}\end{pmatrix}$. Conditioning now on y_i^o, put $U=(y_i^c|Q_i^c, y_i^o)$, $\mu_i^c=E(u)$, $S_i^c=\mathrm{var}(U)$. Then U follows a multinormal distribution $N(\mu_i, S_i)$ left-truncated at Q_i^c. Note that $(y_i|Q_i, C_i)=(y_i|Q_i, C_i, y_i^o)=\mathrm{vec}(y_i^o, (y_i^c|Q_i, y_i^o))$. It follows that

$$E(y_i|Q_i, C_i, \theta)=\mathrm{vec}(y_i^o, \mu_i^c), \mathrm{var}(y_i|Q_i, C_i, \theta)=\begin{pmatrix}0 & 0\\ 0 & S_i^c\end{pmatrix}.$$

ii. The second step is to transform U to a truncated unit multinormal variable, for which closed-form formulas are available. Let B_i be a diagonal matrix with diagonal equal to the square roots of the corresponding diagonal elements in S_i. Put $T=B_i^{-1}(U-\mu_i)$. Then T has a multinormal distribution $N(0, R_i)$ left-truncated at $a_i=B_i^{-1}(Q_i^c-\mu_i)$, and $R_i=B_i^{-1}S_iB_i^{-1}$ is the correlation matrix corresponding to S_i. Then $\mu_i^c=B_iE(T)+\mu_i$, $S_i^c=B_i\mathrm{var}(T)B_i$, and calculation of μ_i^c, S_i^c reduces to computing the mean and variance of T.

iii. Finally, formulas for $E(T)$, var(T) were developed by Tallis (1961) and Finney (1962), and their derivation and computation are given in Appendix A3. These formulas are available in closed form, depending on the multinormal CDF. The latter is available in R through the `pmvnorm()` function from the `mvtnorm` package (Genz 1992; Genz, Bretz, and Hothorn 2006). Note that except for the one- and two-dimensional cases, `pmvnorm()` uses internally a randomized quasi-Monte Carlo algorithm; see also Genz (1993, 2008).

The variance of the MLE $\hat{\theta}$, estimated at convergence, is adjusted for the censored information using Louis's formula (Orchard and Woodbury 1972; Louis 1982). The variance of the fixed effects in the approximate MLE is given (Hughes 1999) by

$$\mathrm{var}(\hat{\beta})=\left(\sum_{i=1}^{m}\{X_i^\top V_i^{-1}X_i-X_i^\top V_i^{-1}\mathrm{var}(y_i|Q_i,C_i)V_i^{-1}X_i\}\right)^{-1}. \quad (2.8)$$

2.2 Restricted Maximum Likelihood Estimation

For the computation of REML (Harville 1977; Bates and Pinheiro 1997) the M-step formulas for the variance components, (2.6)–(2.7) need appropriate adjustments. Put $\xi = (\sigma^2, F)$, so that $\theta = (\beta, \xi)$. The formula for $\hat{\delta}$ is obtained in the same way as in the MLE case (2.5):

$$\hat{\delta}=(M^\top M)^{-1}M^\top E(\tilde{y}|Q,C,\beta,\xi). \quad (2.9)$$

In contrast, in the remainder of the section the conditional expectations are conditional on ξ, Q, and C, but not on β, which is integrated out, assuming the improper distribution $p(\beta) \sim 1$. A competing estimator to (2.9) is

$$\hat{\delta}_R=(M^\top M)^{-1}M^\top E(\tilde{y}|Q,C,\xi). \quad (2.10)$$

For the variance components, the M-step updates are obtained similarly to the MLE case, except that β is treated as a random variable in the likelihood; see Appendix A3 for details. We update σ^2 from the EM with only $\{y_{ij} : C_{ij} = 1\}$ as missing data, and Ψ from the EM with $\{y_{ij} : C_{ij} = 1\}$, b, and β as missing data. The resulting equations are

$$\hat{\sigma}^2_R=\frac{1}{n-p}[\,\|E(\tilde{y}|Q,C,\xi)-M\hat{\delta}_R\|^2+\mathrm{tr}(\{I-M(M^\top M)^{-1}M^\top\}\mathrm{var}(\tilde{y}|Q,C,\xi))]. \quad (2.11)$$

$$\hat{\Psi}_R=\frac{1}{m}\sum_{i=1}^{m}E(b_ib_i^\top|Q,C,\xi)=\frac{1}{m}\sum_{i=1}^{m}E(b_i|Q,C,\xi)E(b_i|Q,C,\xi)^\top+\frac{1}{m}\sum_{i=1}^{m}\mathrm{var}(b_i|Q,C,\xi) \quad (2.12)$$

for the unstructured case; p is the dimension of β. The conditional mean and variance for b are computed as for the LME case, using that $\delta|y \sim N(\hat{\delta}, \hat{\sigma}^2(M^\top M)^{-1})$. These expressions are

$$E(\delta|Q,C,\xi)=\hat{\delta}_R=(M^\top M)^{-1}M^\top E(\tilde{y}|Q,C,\xi),$$
$$\mathrm{var}(\delta|Q,C,\xi)=\sigma^2(M^\top M)^{-1}+(M^\top M)^{-1}M^\top\mathrm{var}(\tilde{y}|Q,C,\xi)M(M^\top M)^{-1}.$$

The final step is determining $E(y|Q, C, \xi)$, $\mathrm{var}(y|Q, C, \xi)$. The distribution of $(y|Q, C, \xi)$ is that of $(y|\xi)$, truncated to the region $\mathscr{R} = \{y : y_{ij} \leq Q_{ij}$ for $C_{ij} = 1$, $y_{ij} = Q_{ij}$ for $C_{ij} = 0$, $i = 1, \ldots, m, j = 1, \ldots, n_i\}$:

$$p(y|Q,C,\xi)\propto p(Q,C|y,\xi)p(y|\xi)=p(y|\xi)I[y\in\mathscr{R}].$$

However, $(y|\xi)$ does not have a proper distribution. A suitable approximation is given by $E(y|Q, C, \xi) \approx E(y|Q, C, \theta)$, $\mathrm{var}(y|Q, C, \xi) \approx \mathrm{var}(y|Q, C, \theta)$.

2.3 The Likelihood Function

Let $\Phi_n(u; A)$ and $\varphi(u; A)$ be respectively the left-tail probability (component-wise) and the probability density function of the $N(0, A)$ distribution, computed at u. Let $\alpha_i=P(y_i^c<Q_i^c|y_i^o)=\Phi_{n_i^c}(a_i;R_i)$. The likelihood for cluster i is given by

$$L_i=P(Q_i|C_i,\theta)=P(y_i^c\leq Q_i^c|y_i^o=Q_i^o,\theta)P(y_i^o=Q_i^o|\theta)=\alpha_i\phi(Q_i^o-X_i^o\beta;\Sigma_{oo}).$$

Therefore, the log-likelihood function for the observed data is given by

$$l(\theta)=\sum_{i=1}^{m}\{\log\alpha_i+\log\phi_{n_i^o}(Q_i^o-X_i^o\beta;\Sigma_{oo})\}. \quad (2.13)$$

This can be computed at each step of the EM algorithm without additional computational burden, because α_i's are computed at the E-step (see Appendix A3). The log-likelihood can be used to monitor the convergence of the algorithm. Alternatively, Vaida, Fitzgerald, and DeGruttola (2007) monitored convergence using the objective function

$$f_o(\theta)=-n/2\{1+\log(2\pi\sigma^2)\}+m/2\log|F^{-1}|-1/2\sum_{i=1}^{m}\log|Z_i^\top Z_i+F^{-1}|. \quad (2.14)$$

which is the log-likelihood of the linear mixed model without censoring, with β profiled out (Pinheiro and Bates 2000, chap. 2). An analogous objective function is defined for REML estimation. Model selection based on the observed likelihood can be done using appropriate likelihood ratio tests, AIC, and BIC.

Convergence is declared when the improvement in log-likelihood falls below a certain preset limit. In practice, `pmvnorm` shows small random variability, which leads to nonincreasing log-likelihood beyond a certain level. At that point the algorithm has reached its limit of precision, and it can be stopped. The variability due to `pmvnorm` can be controlled using the `algorithm=GenzBretz` (`value`) argument.

3. MORE GENERAL LINEAR MIXED EFFECTS MODELS

The methodology from the previous section applies to a general linear mixed effects model (2.1). The formulas (2.5) and (2.6) apply unchanged when $\Lambda = I_n$. For the variance components, the M-step formula depends on the specific structure of D, and it is the solution of the general equation (A6) in the Appendix. Section 3.2 below shows a general method of dealing with the estimation of Λ. In this section we make explicit the EM derivation for several LME of practical importance.

3.1 Variance Matrices for the Random Effects

Following Pinheiro and Bates (2000), we will consider several structures for Ψ. The M-step updates are as follows (see Appendix A1 for details):

1. *Unstructured* Ψ. $\hat{\Psi}=A=m^{-1}\sum_{i=1}^{m}E(b_i b_i^{\top}|Q,C,\theta^*)$, which is (2.7).
2. *Diagonal* Ψ. $\hat{\Psi}=\mathrm{diag}(m^{-1}\sum_{i=1}^{m}E(b_{ik}^2|Q,C,\theta^*);k=1,\ldots,q)$.
3. *Block-diagonal* Ψ. Let $\Psi = \mathrm{diag}(\Psi_1, \ldots, \Psi_K)$, and let $A_1, \ldots, A_K$ be the corresponding block-diagonal submatrices of A. Then $\hat{\Psi} = \mathrm{diag}(A_1, \ldots, A_K)$.
4. *Multiple of identity*. Let $\Psi = \tau^2 I$. Then $\hat{\tau}^2=\frac{1}{mq}\sum_{i=1}^{m}\sum_{k=1}^{q}E(b_{ik}^2|Q,C)$.
5. *Compound symmetry*. Assume $\Psi = \tau^2 I + \gamma^2 J$, where J is the square matrix with all entries equal to 1. Then $\hat{\Psi}=\frac{\mathrm{tr}(A)}{q}I+\frac{\mathrm{sum}(A)-\mathrm{tr}(A)}{q(q-1)}J$, where $\mathrm{sum}(A) = \sum_{i,j} A_{ij}$.

3.2 Heteroscedastic Error

Consider the extended linear mixed effects model

$$y_i=X_i\beta+Z_i b_i+e_i, b_i\sim N(0,\sigma^2 F), e_i\sim N(0,\sigma^2\Lambda_i), i=1,\ldots,m, \quad (3.1)$$

where the Λ_i are positive-definite matrices parameterized by a small set of parameters λ. Let $\Lambda_i^{1/2}$ be the square root of Λ_i, such that $\Lambda_i=(\Lambda_i^{1/2})^{\top}\Lambda_i^{1/2}$, and let $\Lambda_i^{1/2}$ be its inverse. Further, define

$$y_i^*=(\Lambda_i^{-1/2})^\top y_i, \quad e_i^*=(\Lambda_i^{-1/2})^\top e_i,$$
$$X_i^*=(\Lambda_i^{-1/2})^\top X_i, \quad Z_i^*=(\Lambda_i^{-1/2})^\top Z_i.$$

Then

$$y_i^*=X_i^*\beta+Z_i^*b_i+e_i^*, b_i\sim N(0,\sigma^2F), e_i^*\sim N(0,\sigma^2I), i=1,\ldots,m. \quad (3.2)$$

That is, given the parameter λ, y_i^* is described by a basic linear mixed effects model, and the parameters β, σ^2, Ψ can be estimated using (2.5)–(2.7).

To estimate λ, it is convenient to write $\Lambda_i = V_iK_iV_i$, where V_i is a diagonal matrix with elements $\sqrt{\mathrm{var}(e_{ij})}/\sigma$ on the diagonal, and $K_i = \mathrm{Corr}(e_i)$. The parameters in V_i and K_i are assumed independent, that is, $\lambda = \mathrm{vec}(\lambda_\upsilon, \lambda_\kappa)$, and $V_i = V_i(\lambda_\upsilon)$, $K_i = K_i(\lambda_\kappa)$. Then λ_υ and λ_κ can be estimated separately, as discussed below, and they depend on the models assumed for the variance function and within-subject correlation structure, respectively.

As in Pinheiro and Bates (2000, chap. 5), we consider the following cases for the variance function. See Appendix A2 for details.

1. *Fixed variance*: $\mathrm{var}(e_{ij}) = \sigma^2\upsilon_{ij}$, where υ_{ij} is fixed; alternatively, $\upsilon_{ij} = \upsilon_{ij}(\mu_{ij})$, $\mu_{ij} = X_{ij}\beta + Z_{ij}b_i$.

   For fixed υ_{ij}, Λ_i is known, and the estimation proceeds based on the transformed model (3.2). For $\upsilon_{ij} = \upsilon_{ij}(\mu_{ij})$, an approximation is used (Davidian and Giltinan 1995), with μ_{ij} and υ_{ij} computed at each step using the current parameter values, followed by the parameter update using model (3.2).

2. *Different variances per stratum*: $\mathrm{var}(e_{ij}) = \sigma^2\delta_k$, where $k = k(i, j)$ corresponds to the stratum, $k = 1, \ldots, K$.

   After reordering the error terms e_{ij} according to the variance strata, we can write

   $$\Lambda=\sigma^{-2}\mathrm{var}(e)=\mathrm{diag}(\delta_1 I_{k_1},\ldots,\delta_s I_{k_s}), \quad (3.3)$$

   where s is the number of strata. For identifiability, we assume that $\delta_1 = 1$. Then, the M-step update is

   $$\delta_j=\frac{1}{k_j\sigma^2}\mathrm{tr}E(e_je_j^\top|Q,C), j=2,\ldots,K,$$

   where e_j is the subvector of e corresponding to δ_j; see Appendix A2 for details. Put $U = Z(D^{-1} + Z^\top\Lambda^{-1}Z)^{-1}Z^\top$. Further calculations reveal that

$$
\begin{aligned}
&E(ee^\top|Q,C)\\
&\quad=E(e|Q,C)E(e|Q,C)^\top\\
&\quad+\mathrm{var}(e|Q,C)\\
&\quad=(I-U\Lambda^{-1})(E(y|Q,C)\\
&\qquad-X\beta)(E(y|Q,C)-X\beta)^\top(I-U\Lambda^{-1})^\top\\
&\quad+(I-U\Lambda^{-1})\mathrm{var}(y|Q,C)(I-U\Lambda^{-1})^\top+U.
\end{aligned}
\tag{3.4}
$$

3. *Exponential of covariate*: $\mathrm{var}(e_{ij}) = \sigma^2\exp(2\upsilon_{ij}\delta)$, with υ_{ij} fixed, or depending on μ_{ij}. Then δ is the solution of the equation

$$\mathrm{tr}(\{I-\sigma^{-2}\Lambda^{-1}E(ee^\top|Q,C)\}\mathrm{diag}(2\upsilon_{ij}))=0.$$

which can be solved using a generic equation solver; $E(ee^\top|Q, C)$ is given by (3.4).

4. *Power of covariate*: $\mathrm{var}(e_{ij}) = \sigma^2|\upsilon_{ij}|^{2\delta}$, where υ_{ij} is either fixed (e.g., a covariate), or depends on μ_{ij}.

   This reduces to the case above, seeing that $|\upsilon_{ij}|^{2\delta} = \exp(2w_{ij}\delta)$, with $w_{ij} = \log(|\upsilon_{ij}|)$.

3.3 Autocorrelated Error

The proposed framework allows more general dependence structures (i.e., Λ not diagonal). For example, in a *first-order autoregressive model* (AR(1)), $\mathrm{var}(e_i) = \sigma^2\Lambda_i$, with the (k, l) element of Λ_i given by $\rho_{kl} = \rho^{|k-l|}$; a more general model is $\rho_{ikl} = \rho^{d(t_{ik}, t_{il})}$, where t_{ij} is a known covariate. The parameter ρ is updated as the solution of the equation

$$\mathrm{tr}(\{\Lambda^{-1}-\sigma^{-2}\Lambda^{-1}E(ee^\top|Q,C)\Lambda^{-1}\}\dot{\Lambda})=0.$$

where $\dot{\Lambda} = \mathrm{diag}(\dot{\Lambda}_1, \ldots, \dot{\Lambda}_m)$, and the generic (k, l) element of $\dot{\Lambda}_i$ is $d(t_{ik}, t_{il})\rho^{d(t_{ik}, t_{il})-1}$. As before, $E(ee^\top|Q, C)$ is given by (3.4). The *continuous-time* AR(1) *model* obtains for $d(t_{ik}, t_{il}) = |t_{ik} - t_{il}|$. See Appendix A2 for details.

3.4 Multilevel Models

The N/LMEC methodology extends in a natural way to multilevel models. For example, in the notation of Pinheiro and Bates (2000, chap. 2), consider the nested model

$$y_{ij}=X_{ij}\beta+Z_{i,j}b_i+Z_{ij}b_{ij}+e_{ij},\ i=1,\ldots,m, j=1,\ldots,m_i, b_i\sim N(0,\Psi_1), b_{ij}\sim N(0,\Psi_2), e_{ij}\sim N(0,\sigma^2 I). \tag{3.5}$$

We assume that b_i, b_{ij}, e_{ij} are all independent for different i, j. Let n_{ij} be the size of the vector y_{ij}, and $n = \sum_{i,j} n_{ij}$. As before, let Q be the vector of observed values y or censoring limits, and C the censoring indicator.

Put $b_{i0} = \mathrm{vec}(b_i, b_{i1}, \ldots, b_{im_i})$ for each i; $Z_i = (\mathrm{vec}(Z_{i,1}, \ldots, Z_{i, m_i})\ \mathrm{diag}(Z_{i1}, \ldots, Z_{im_i}))$; similarly, $X_i = \mathrm{vec}(X_{i1}, \ldots, X_{im_i})$; $y_i = \mathrm{vec}(y_{i1}, \ldots, y_{im_i})$; $e_i = \mathrm{vec}(e_{i1}, \ldots, e_{im_i})$. Also, $y =$

vec(y_1, …, y_m), X = vec(X_1, …, X_m), Z = vec(Z_1, …, Z_m), e = vec(e_1, …, e_m), b = vec(b_1, …, b_m). Then (3.5) becomes

$$y_i=X_i\beta+Z_ib_i+e_i, b_i\sim N(0,\Psi_{i0}), e_i\sim N(0,\sigma^2I), i=1,\ldots,m,$$

which is identical to the Laird–Ware model (2.2), except for the fact that Ψ_{10}, …, Ψ_{m0} are not necessarily of same dimension: Ψ_{i0} = diag(Ψ_1, diag$_{m_i}$ (Ψ_2)). Put $F_i=\sigma^{-2}\Psi_{i0}=(\Delta_{i0}^\top\Delta_{i0})^{-1}$. The formulas for $\hat{\delta}$ and $\hat{\sigma^2}$ are derived analogously to (2.5) and (2.6):

$$\hat{\delta}=(M^\top M)^{-1}M^\top E(\tilde{y}|Q,C), \quad (3.6)$$

$$\hat{\sigma}^2=\frac{1}{n}\|E(\tilde{y}|Q,C)-M\hat{\delta}\|^2+\frac{1}{n}\sum_{i=1}^m\mathrm{tr}\{\mathrm{var}(y_i|Q,C)\}-\frac{1}{n}\sum_{i=1}^m\mathrm{tr}\{W_iZ_i^\top\mathrm{var}(y_i|Q,C)Z_i\}, \quad (3.7)$$

with $W_i=(Z_i^\top Z_i+F_{i0}^{-1})^{-1}$.

We show in Appendix A2 that the M-step for Ψ_1, Ψ_2, in the unstructured case, is given by

$$\hat{\Psi}_1=\frac{1}{m}\sum_{i=1}^m E(b_ib_i^\top|Q,C), \hat{\Psi}_2=\frac{1}{\sum_i m_i}\sum_{i,j}E(b_{ij}b_{ij}^\top|Q,C).$$

The conditional expectations above can be computed using the expressions

$$E(b_{i0}|Q,C)=W_iZ_i^\top\{E(y_i|Q,C)-X_i\beta\}, \mathrm{var}(b_{i0}|Q,C)=\sigma^2W_i+W_iZ_i^\top\mathrm{var}(y_i|Q,C)Z_iW_i.$$

4. NONLINEAR CASE

The NLME (Lindstrom and Bates 1990; Pinheiro and Bates 2000) is given by

$$y_{ij}=f(\beta,b_i)+e_{ij}, \quad (4.1)$$

where $f(\beta, b_i) = f(\beta, b_i, x_{ij})$ is a nonlinear function of the fixed β and random effect b_i ; x_{ij} is a vector of covariates, and b_i and e_{ij} are given by (2.2). The approximate MLE ($\hat{\beta}$, $\hat{\sigma^2}$, $\hat{F}$) and predictors for the random effects $\hat{b_i}$ are computed by iterative linearization (L) of the conditional mean function. The L-step yields the LME

$$w_i=X_i^*\beta+Z_i^*b_i+e_i, i=1,\ldots,m, \text{where } w_i=y_i-\{f_i^*-X_i^*\beta^*-Z_i^*b_i^*\}, \quad (4.2)$$

$X_i^*=\frac{\partial f_i^*}{\partial\beta}$, $Z_i^*=\frac{\partial f_i^*}{\partial b_i}$, y_i is the n_i-vector dependent variable for the ith subject, f_i, e_i are respectively the corresponding mean function and error n_i-vectors, and the starred terms are computed at the current parameters (β^*, b_i^*). For censored response the linearized model (4.2) is an LME with censored data, with same structure as (2.2), which is then solved as indicated in the previous section. The model matrix for (4.2) depends on the current

parameter value, and needs to be recalculated at each iteration. The algorithm iterates to convergence between L-, E-, and M-steps. See Vaida, Fitzgerald, and DeGruttola (2007) for details.

5. CASE STUDIES

We illustrate the proposed method with the analysis of two HIV datasets. The analysis was performed using the R package `lmec` (Vaida and Liu 2009) available on CRAN and `nlmec` available in the online supplement.

5.1 HIV-1 Viral Load After Unstructured Treatment Interruption

The first application concerns a study of 72 perinatally HIV-infected children and adolescents (Saitoh et al. 2008). Unstructured treatment interruption (UTI) is common in this population, due mainly to treatment fatigue. Suboptimal adherence can lead to antiretroviral resistance and diminished treatment options in the future. The subjects in the study had taken antiretroviral therapy for at least 6 months before UTI, and the medication was discontinued for more than 3 months. The HIV viral load from the closest time points at 0, 1, 3, 6, 9, 12, 18, 24 months after UTI were studied. The number of observations from baseline (month 0) to month 24 are 71, 62, 58, 57, 43, 34, 24, and 13, respectively. Out of 362 observations, 26 (7%) observations were below the detection limits, 50 or 400 copies/mL, and were censored at these values. The individual profiles of viral load at different follow-up times after UTI appear in Figure 1. A profile LME model with random intercepts b_i is considered:

$$y_{ij}=b_i+\beta_j+e_{ij}, \quad (5.1)$$

where y_{ij} is the $\log_{10}$ HIV RNA for subject i at time t_j, $t_0 = 0$, $t_1 = 1$, $t_2 = 3$, $t_3 = 6$, $t_4 = 9$, $t_5 = 12$, $t_6 = 18$, $t_7 = 24$ months. One interesting feature of the study is that the probability of dropout depends on the observed viral load outcome. Indeed, by 12 months half of the subjects remained in the study, as subjects went back on treatment when their viral load increased. Assuming that the dropout probability is adequately modeled as a function of the observed trajectory y_i (missing at random; see, e.g., Diggle et al. 2002, chap. 13), this model also adequately deals with the informative dropout. Conditionally on model (5.1) being correct, the dropout does not bias the inference regarding the mean values β_j. This is not the case if the longitudinal structure is ignored, and inference at each time point is based only on data observed at that time point. The mean viral load $E(y_{ij}) = \beta_j$ increased gradually throughout 24 months from 3.62 $\log_{10}$RNA at the time of UTI with 95% CI (3.37–3.87) to 4.38 with 95% CI (4.11–4.64) at 6 months, then to 4.58 (95% CI 4.29–4.88) at 12 months and 4.81 (95% CI 4.41–5.21) at 24 months. This is in contrast with the mean profiles of the observed data alone, which show a leveling off and a decrease in viral load between 6 and 12 months (see Figure 1). The between-subjects (b_i) and within-subjects (e_{ij}) standard deviations were 0.88 and 0.58 $\log_{10}$, respectively.

5.2 HIV-1 Viral Load Setpoint for Acutely Infected Subjects

The second AIDS case study concerns 320 untreated individuals with acute HIV infection from the AIEDRP Program, a large multicenter observational study of subjects with acute

and early HIV infection. During the acute stage of infection, the large HIV RNA observations may lie *above* the limit of quantification of the assay, which we treat as right-censoring. This limit is between 75,000 and 500,000 cp/mL, depending on the assay. The time of infection was estimated at 24 days prior to first positive HIV RNA sample or detectable serum p24 antigen test. We included the HIV RNA observations in the first 180 days of follow-up and only up to the start of antiretroviral treatment. The subjects had between one and 14 observations: 129 had one, 82 had two, and 109 had three or more observations. Of the 830 recorded observations, 185 (22%) were above the limit of quantification of the assay (right-censored). In the absence of treatment, following acute infection the HIV RNA decreases and then varies around a setpoint value. This setpoint value may differ between individuals, and is of central interest here. The viral setpoint characterizes the severity of infection, it may relate to the strength of the subject's immune system, and it may predict clinical progression of the disease. The individual profiles and a smooth mean of the observed data are included in Figure 2. The smooth curve agrees qualitatively with the postulated shape of the HIV RNA trajectory for acutely infected patients. There is possible indication of a continuing viral decay rather than stabilization to a setpoint, with the caveat that the observed mean curve may be biased due to the exclusion of the censored values and to differential follow-up (see, e.g., Diggle et al. 2002, chap. 11). It is clear that the viral setpoint values differ from subject to subject.

Our analysis considers three models for these data. We started by fitting a four-parameter logistic model taking into account the censoring information. The model is

$$y_{ij}=\alpha_{1i}+\alpha_2[1+\exp\{(t_{ij}-\alpha_3)/\alpha_4\}]^{-1}+e_{ij}. \quad (5.2)$$

where y_{ij} is the $\log_{10}$ HIV RNA for subject i at time t_{ij}. This is an inverted S-shaped curve, with the constant value for the later times representing the subject-specific setpoint. The parameters α_{1i} and α_2 are the setpoint value and the decrease from the maximum HIV RNA; α_4 is a scale parameter modeling the rate of decline, and α_3 is a location parameter indicating the time point at which half of the change in HIV RNA is attained. To force the parameters to be positive we reparameterized the model to $\beta_{1i} = \log(\alpha_{1i})$, $\beta_k = \log(\alpha_k)$, $k = 2, 3, 4$. The setpoint α_{1i} was taken to be random: $\beta_{1i}=\beta_1+b_i, b_i\sim N(0,\sigma^2_{b1})$. It is tempting to consider models including random effects for β_3 and β_4, but there are not enough available data in the acute (earliest) phase of infection to allow for inclusion of these random parameters.

The plot of model residuals against time shows a relatively good fit (Figure 3), but it suggests that the model does not capture a time trend in the data after day 50 since infection and an initial increase in viral load (see also Figure 2). In addition, a variogram of the residuals (Figure 4) indicates long-term autocorrelation, which may be due either to bias in modeling the mean term or to genuine serial autocorrelation beyond the random intercept, unaccounted for in the model.

To address the bias concern we added a linear term after day 50 in the second model:

$$y_{ij}=\alpha_{1i}+\alpha_2[1+\exp\{(t_{ij}-\alpha_3)/\alpha_4\}]^{-1}+\alpha_5(t_{ij}-50)+e_{ij}. \quad (5.3)$$

The residuals' plot (Figure 3) indicates a better overall fit, but the variogram shows that the serial autocorrelation is not properly accounted for. This suggests a third model, by adding a *random slope* after day 50:

$$y_{ij}=\alpha_{1i}+\alpha_2[1+\exp\{(t_{ij}-\alpha_3)/\alpha_4\}]^{-1}+\alpha_{5i}(t_{ij}-50)+e_{ij}. \quad (5.4)$$

As in (5.2), we have $\log(\alpha_{1i}) = \beta_{1i} = \beta_1 + b_{1i}$, $\beta_k = \log(\alpha_k)$ for $k = 2, 3, 4$, but $\alpha_{5i} = \beta_5 + b_{5i}$, to allow for increasing HIV RNA trajectories after day 50. Also, (b_{1i}, b_{5i}) are assumed to be iid, multivariate normal with unrestricted variance matrix. The model fit is slightly better than that in the second model, with the smooth mean residual curve in Figure 3 closer to zero, and fitted values between the fitted values of the first two models. More importantly, the residuals show no serial correlation in the variogram (Figure 4).

The results of the analysis are in Table 1. We can use the last model with reasonable confidence for predictions of viral load. For example, at 6 months since infection the average viral load is 4.55 $\log_{10}$ units (in contrast, the setpoint model (5.2) estimates this at 4.83). The individual 6-month viral load estimates vary between 1.63 and 6.65, with 5th and 95th quantiles at 3.37 and 5.50. The average slope after day 50 was negative, $\beta_5 = -0.0035$ $\log_{10}$ HIV/day, with 95% CI (−0.0063, −0.0006). However, the individual slopes α_{5i} included positive values, with 5th and 95th quantiles of −0.0070 and 0.0004.

6. COMPARISON OF CLOSED-FORM AND MONTE CARLO EM ALGORITHMS

We compared the behavior and performance of the closed-form EM algorithm with the MCEM algorithm of Vaida, Fitzgerald, and DeGruttola (2007) via a simulation study with application in both linear and nonlinear mixed effects models. The MCEM algorithm is an improved, faster version of that of Hughes (1999), and it was described in detail by Vaida, Fitzgerald, and DeGruttola (2007). For clarity of comparison, in this simulation the MCEM does not use the closed-form E-step for one or two censored observations per cluster, unlike in the work of Vaida, Fitzgerald, and DeGruttola (2007). To separate the effect of censoring from the effect of N/LME approximation and finite sample bias, we also include results from the complete data (N/LME). The three methods were compared based on average estimate of the parameter, the simulation-based variance, relative bias $E(\hat{\theta} - \theta)/|\theta|$, and coverage probability for 95% confidence intervals $\Pr(\theta \in \hat{I})$. (The parameter of interest θ is estimated by $\hat{\theta}$ and the confidence interval $\hat{I}$.)

6.1 Linear Case

The first simulation, using 4000 simulated datasets, was based on a Cadralazine pharmacokinetics study dataset (Lunn et al. 1999; Vaida and Blanchard 2005). The original dataset consists of plasma drug concentrations from 10 cardiac patients who were given a

single intravenous dose of 30 mg of Cadralazine, an anti-hypertensive drug. Each subject has the plasma drug concentration measured at 2, 4, 6, 8, 10, 24 hr, for a total of six observations per subject. We considered the linear model (6.1) with random intercepts b_{0i} and random slopes b_{1i} which was studied by Vaida and Blanchard (2005):

$$y_{ij}=\beta_0+b_{0i}+(\beta_1+b_{1i})t_{ij}+e_{ij}. \quad (6.1)$$

where

$$(b_{0i}, b_{1i}) \overset{\text{iid}}{\sim} N(0, \Psi), e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

y_{ij} is equal to log(concentration) − log(dose) for the ith subject at time t_{ij}. Parameters in the simulation were chosen similarly to the estimated values from the LME based on the Cadralazine data: $\beta_0 = -2.83$, $\beta_1 = -0.18$, $\sigma = 0.15$, the matrix Ψ has elements $\Psi_{11} = 0.049$, $\Psi_{12} = 0.001$, $\Psi_{22} = 0.002$. Each simulated dataset had 50 subjects, with six observations per subject and a follow-up of 24 hr; 22.8% of all observations were censored and the average maximum number of censoring observations per subject is 3.2. Table 2 presents the average estimates for fixed effects, variance components, and σ; their simulation-based standard deviation; and the numeric error of the estimation. The MLE method was used throughout; the REML results (not included) were similar. Table 3 contains the coverage probabilities of the fixed effects and the relative bias for all estimated parameters. The numeric error of the estimator was computed at the suggestion of one of the referees, as follows: the "gold standard" MLE $\hat{\theta}_G$ was computed using increased precision for `mvtnorm` and for the EM convergence criterion (by two orders of magnitude), at the expense of a much longer running time. This was compared with the "standard" estimator $\hat{\theta}_S$, and the root mean squared error of simulation, $E\{(\hat{\theta}_S - \hat{\theta}_G)^2\}$ was computed for each parameter. The numeric error was defined as the relative root mean squared error, $\text{NErr} = \sqrt{E\{(\hat{\theta}_S - \hat{\theta}_G)^2\}}/|\theta|$.

The closed-form EM,MCEM, and complete data LME performed similarly: all average estimates of parameters were very close to each other, and close to the true values except for Ψ_{12}. These censored data algorithms had slightly larger simulation-based variances than LME, as expected. All three methods underestimated Ψ_{12} with a bias of −78.7% for closed-form EM, −77.2% for MCEM, and −79.5% for LME. Both closed-form EM and MCEM account for censoring in the variance of the parameter estimates. The relative loss of information due to censoring for all six parameters, $1 - \text{var}_{LME}(\hat{\theta})/\text{var}_{EM}(\hat{\theta})$, ranged from 21% for β_0 to 44% for Ψ_{12}. In terms of coverage, all three methods performed well. The coverage probabilities of β_0 were 93.6% for closed-form EM, and 93.3% and 94.3% for MCEM and complete data LME; the coverage probabilities for β_1 were 93.8% for closed-form EM and MCEM, and 94.1% for LME. Note that with 4000 datasets the standard error for the coverage values is 0.4 percentage points. As expected, the closed-form EM improved the computation time substantially. On average, it took 212 sec for MCEM to converge, but only 5.5 sec for closed-form EM, about 40 times faster than MCEM. In addition, the numeric error is practically negligible for the closed-form EM, more than 10 times smaller than MCEM for the mean parameters, and at least twice smaller for the variance components.

The contrast between the closed-form EM and MCEM is illustrated in Figure 5. The figure displays the objective function (2.14), which is the log-likelihood surrogate used by MCEM in the article by Vaida, Fitzgerald, and DeGruttola (2007), for the two algorithms, for the Cadralazine dataset used as a framework for the simulations in this section. The closed-form EM has smooth and fast convergence (12 sec). In contrast, the MCEM log-likelihood is subject to random variations, proportional to the inverse square root of the MCMC sample size, and larger convergence times (120 sec in this example).

6.2 Nonlinear Case

The second simulation was based on the AIEDRP data studied in Section 5.2. We considered a similar logistic model (6.2) with random setpoints α_{1i} and random decline rates α_{4i}:

$$y_{ij}=\alpha_{1i}+\alpha_2[1+\exp\{(t_{ij}-\alpha_3)/\alpha_{4i}\}]^{-1}+e_{ij}. \quad (6.2)$$

Reparameterizing the model (6.2) as in Section 4.2, we estimated $\beta_{si} = \log(\alpha_{si})$ ($s = 1, 4$) and $\beta_t = \log(\alpha_t)$ ($t = 2, 3$) where $\beta_{si} = \beta_s + b_{si}$, $(b_{1i}, b_{4i}) \overset{\text{iid}}{\sim} N(0, \Psi)$, and random errors are normally distributed with a standard deviation of σ. Parameters in the simulation were also chosen similarly to the estimated values based on the original data using LME: $\beta_1 = 1.6094$, $\beta_1 = 0.6931$, $\beta_3 = 3.8067$, $\beta_4 = 2.3026$, $\sigma = 0.55$; the matrix Ψ has elements $\Psi_{11} = 0.0025$, $\Psi_{12} = -0.0010$, $\Psi_{22} = 0.0100$. Each simulated dataset had 100 subjects, with ten observations per subject and a follow-up of 90 days. Seventeen point eight percent of all observations were censored and the average maximum number of censoring observations per subject is 5.1. The simulation is based on 1000 datasets. Table 4 presents the average estimates for fixed effects, variance components, and σ, and their simulation-based standard deviations. Table 5 contains the coverage probabilities of the fixed effects and the relative bias for all estimated parameters.

The closed-form EM and MCEM still performed very similarly; the average estimates of all parameters were very close. Compared to NLME, they also provided similar estimates for fixed effects and σ. All three methods overestimated Ψ_{11} and Ψ_{22} and underestimated Ψ_{12}. The bias for closed-form EM and MCEM is 15.96% and 16.19% for Ψ_{11}, −83.89% and −84.79% for Ψ_{12}, and 392% and 431% for Ψ_{22}. The relative loss of information due to censoring ranged from 4% to 23% for fixed effects and σ. For variance components, the relative loss is 0% for Ψ_{11}, 27% for Ψ_{12}. For Ψ_{22}, the closed-form EM had a smaller variance than NLME; this might be due to more nonconvergent iterations in the simulation for NLME. The coverage probabilities for fixed effects β_i ($i = 1, 2, 3, 4$) using the closed-form EM and MCEM are similar and all higher than NLME. The range of the coverage probabilities was 92.2%–94.6% for closed-form EM, 92.1%–94.5% for MCEM, and 71.4%–93.5% for NLME. In terms of computation time, the closed-form EM is 5 times faster than the MCEM.

7. DISCUSSION

In this article we have developed an EM algorithm for linear and nonlinear mixed effects models with censored response. The algorithm has a closed-form expression for the E-step, based on formulas for the mean and variance of the truncated multinormal distribution. The computation uses existing functions for the multinormal cumulative distribution function. Our simulation studies showed that this implementation leads to an improved speed of convergence of up to an order of magnitude over leading alternatives. The observed likelihood is derived at no additional computational cost, paving the way for model selection procedures, such as the likelihood ratio test, the AIC, or the conditional AIC (Vaida and Blanchard 2005). As an added benefit, the EM likelihood sequence is monotonic and the difficulties in assessing convergence which face MCEM algorithms are avoided.

We considered several extensions of the standard Laird–Ware model, including heteroscedastic and autocorrelated error, and multilevel models. Other cases, such as crossed random effects (Breslow and Clayton 1993; Vaida and Meng 2005, see, e.g.) are not explicitly discussed, but are covered by the general methodology presented here. For these, however, the updates of the variance components need to be made on a case-by-case basis.

Finally, it is worth emphasizing that we are currently using these methods in our biostatistical practice. The articles by Saitoh et al. (2008) and Cysique et al. (2009) are two such instances of ‘real’ data analysis in the medical literature.

ACKNOWLEDGMENTS

Florin Vaida acknowledges partial support from NIH grants AI-43638, MH-62512, MH-22005, and AI-47033. Lin Liu’s work was supported in part by NIH grants AI-43638 and AI-074621. We are grateful to the editor, associate editor, and three anonymous referees for their suggestions which greatly improved the paper.

REFERENCES

Bates, DM.; Pinheiro, JC. technical report. Murray Hill, NJ: Bell Labs; 1997. Computational Methods for Multilevel Modelling.

Bates D, Sarkar D. lme4: Linear Mixed-Effects Models Using S4 Classes. R Package Version. 2007:0.9975–0.9913.

Breslow N, Clayton D. Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association. 1993; 88:9–25.

Cysique LA, Vaida F, Letendre S, Gibson S, Cherner M, Woods SP, McCutchan JA, Heaton RK, Ellis RJ. Dynamics of Cognitive Change in Impaired HIV-Positive Patients Initiating Antiretroviral Therapy. Neurology. 2009; 73:342–348. [PubMed: 19474412]

Davidian, M.; Giltinan, DM. Nonlinear Models for Repeated Measurement Data. Boca Raton, FL: Chapman & Hall; 1995.

Diggle, PJ.; Heagerty, P.; Liang, K-Y.; Zeger, SL. Analysis of Longitudinal Data. 2nd ed.. Oxford, England: Oxford University Press; 2002.

Finney DJ. Cummulants of Truncated Multinormal Distribution. Journal of the Royal Statistical Society, Ser. B. 1962; 24:535–536.

Genz A. Numerical Computation of Multivariate Normal Probabilities. Journal of Computational and Graphical Statistics. 1992; 1:141–150.

Genz A. Comparison of Methods for the Computation of Multivariate Normal Probabilities. Computing Science and Statistics. 1993; 25:400–405.

Genz, A. MCQMC Methods for Multivariate Statistical Distributions. In: Keller, A.; Heinrich, S.; Niederreiter, H., editors. Monte Carlo and Quasi-Monte Carlo Methods 2006. New York: Springer; 2008. p. 35-52.

Genz A, Bretz F, Hothorn T. mvtnorm: Multivariate Normal and *t* Distribution. R Package Version 0.7-5. 2006

Harville DA. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. Journal of the American Statistical Association. 1977; 72:320–340.

Hughes JP. Mixed Effects Models With Censored Data With Applications to HIV RNA Levels. Biometrics. 1999; 55:625–629. [PubMed: 11318225]

Jiang, J. Generalized Linear Mixed Models: Theory and Applications. New York: Springer; 2007.

Lindstrom MJ, Bates DM. Nonlinear Mixed Effects Models for Repeated Measures Data. Biometrics. 1990; 46:673–687. [PubMed: 2242409]

Louis TA. Finding the Observed Information Matrix When Using the EM Algorithm. Journal of the Royal Statistical Society, Ser. B. 1982; 44:226–233.

Lunn, D.; Wakefield, J.; Andrew, T.; Best, N.; Spiegelhalter, D. PKBugs Users Guide. London: Imperial College of Science, Technology and Medicine; 1999.

Magnus, JR.; Neudecker, H. Matrix Differential Calculus With Applications in Statistics and Econometrics. Chichester: Wiley; 1999.

Orchard, T.; Woodbury, MA. Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability. Vol. Vol. 1. Berkeley, CA: University of California Press; 1972. A Missing Information Principle, Theory and Application; p. 679-715.

Pinheiro, JC.; Bates, DM. Mixed-Effects Models in S and S-PLUS. New York: Springer; 2000.

Pinheiro J, Bates D, DebRoy S, Sarkar D. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-78. 2006

Ruppert, D. Comment on “Maximization by Parts in Likelihood Inference”. In: Song, PX-K.; Fan, Y.; Kalbfleisch, JD., editors. Journal of the American Statistical Association. Vol. 100. 2005. p. 1161-1163.

Saitoh A, Foca M, Viani RM, Heffernan-Vaca S, Vaida F, Lujan-Zilbermann J, Emmanuel PJ, Deville JG, Spector SA. Clinical Outcome in Perinatally Acquired HIV-Infected Children and Adolescents After Unstructured Treatment Interruption. Pediatrics. 2008; 121:e513–e521. [PubMed: 18310171]

Tallis GM. The Moment Generating Function of the Truncated Multi-Normal Distribution. Journal of the Royal Statistical Society, Ser. B. 1961; 23:223–229.

Vaida F, Blanchard S. Conditional Akaike Information for Mixed Effects Models. Biometrika. 2005; 92(2):351–370.

Vaida F, Liu L. lmec: Linear Mixed-Effects Models With Censored Responses. R Package Version 1.0. 2009

Vaida F, Meng X-L. Two Slice-EM Algorithms for Fitting Generalized Linear Mixed Models With Binary Response. Statistical Modelling. 2005; 5:229–242.

Vaida F, Fitzgerald A, DeGruttola V. Efficient Hybrid EM for Nonlinear Mixed Effects Models With Censored Response. Computational Statistics and Data Analysis. 2007; 51:5718–5730. [PubMed: 19578533]

Wand MP. Vector Differential Calculus in Statistics. The American Statistician. 2002; 56:55–62.

Wolfinger, R. Technical Report 287. Cary, NC: SAS Institute Inc.; 1999. Fitting Nonlinear Mixed Models With the New NLMIXED Procedure.

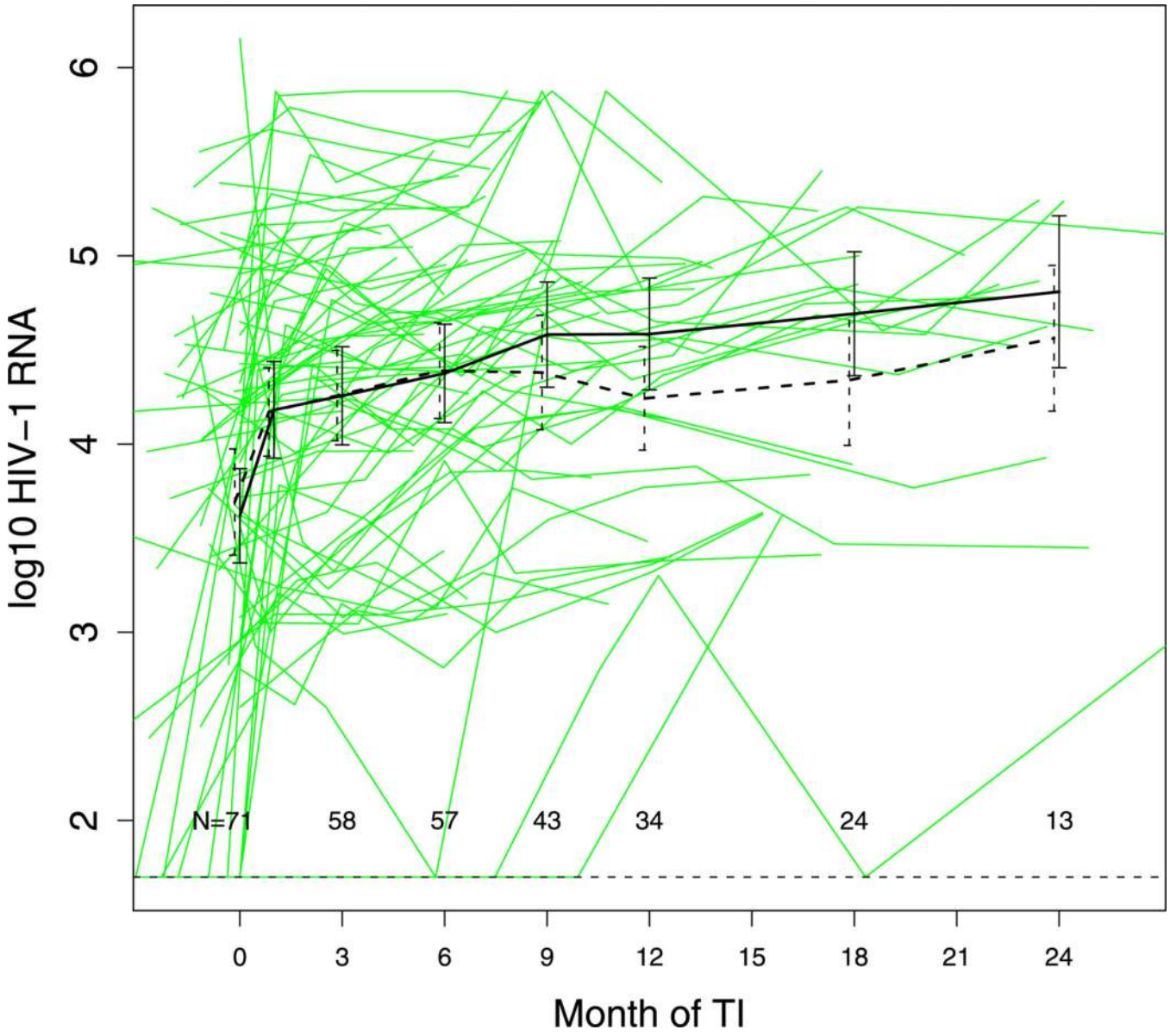

Figure 1.

UTI data: Individual profiles and overall mean (and 95% CI) log10 viral load at different follow-up times post-UTI. The means are estimated (i) using a random intercept model, with adjustment for viral load values censored below the limit of detection (–); (ii) using observed data alone (- -).

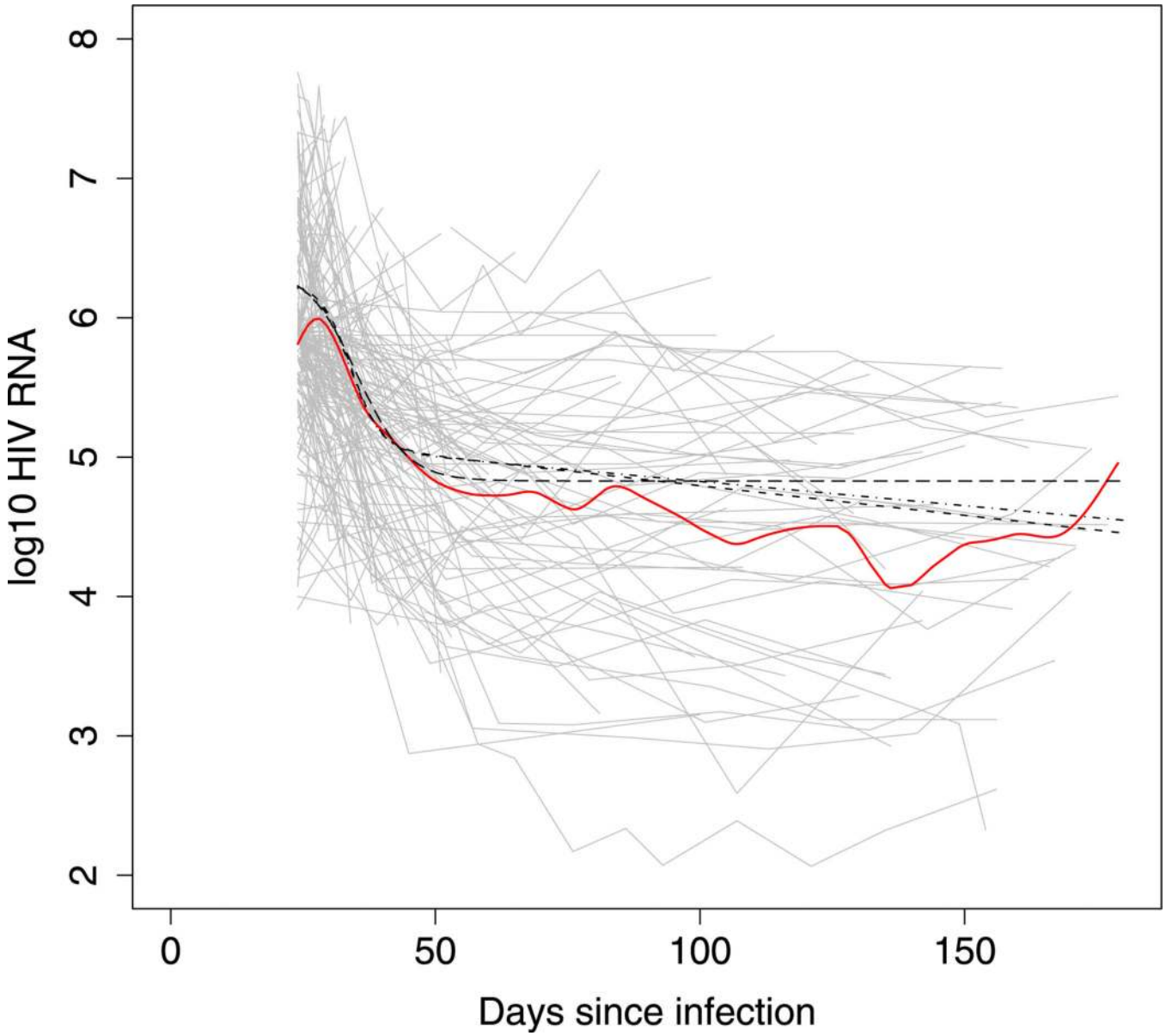

Figure 2.

AIEDRP data: Model fits from (i) random intercept logistic model (– –); (ii) random intercept logistic model with linear decrease after 50 days (- -); (iii) logistic model with random intercept and random linear decrease after 50 days (–·–). Solid line: a smooth fit of the observed data, with censored observations excluded.

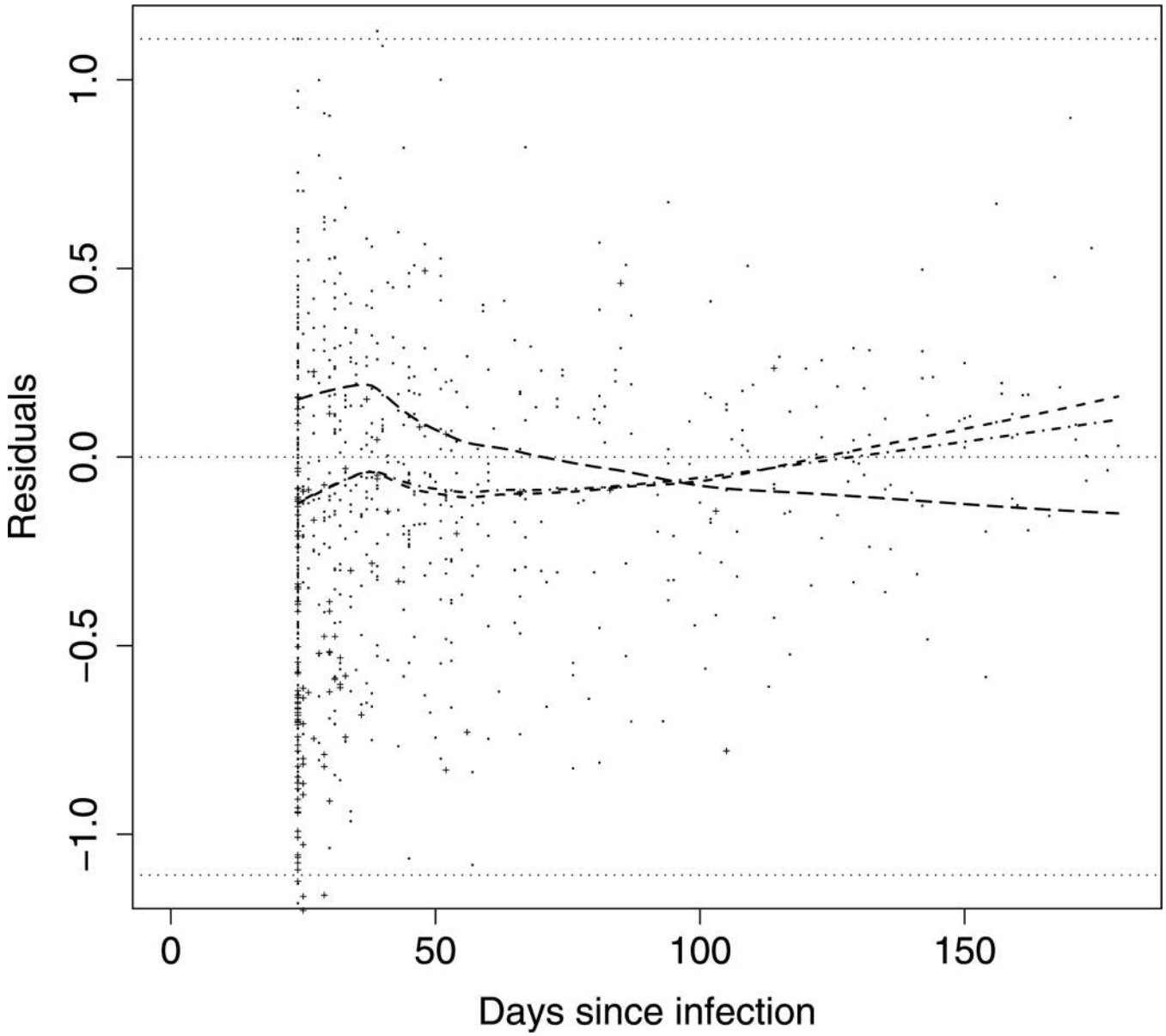

Figure 3.

AIEDRP data: Smooth means of residuals from (i) random intercept logistic model (– –); (ii) random intercept logistic model with linear decrease after 50 days (- -); (iii) logistic model with random intercept and random linear decrease after 50 days (–·–). The residuals from model (iii) appear as points; the right-censored residuals appear as "+."

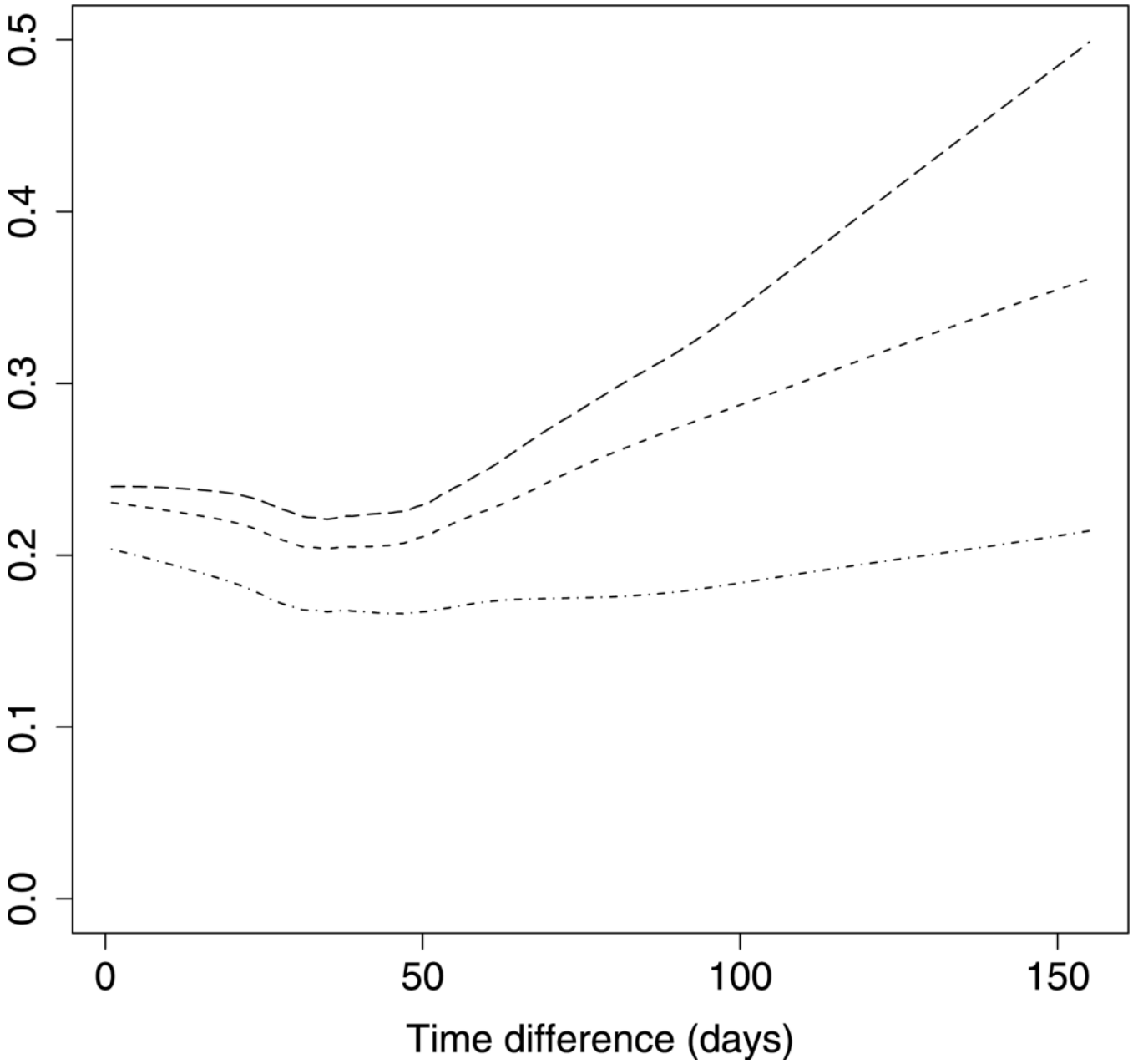

Figure 4.

AIEDRP data: Variogram from model residuals from (i) random intercept logistic model (– –); (ii) random intercept logistic model with linear decrease after 50 days (- -); (iii) logistic model with random intercept and random linear decrease after 50 days (–·–).

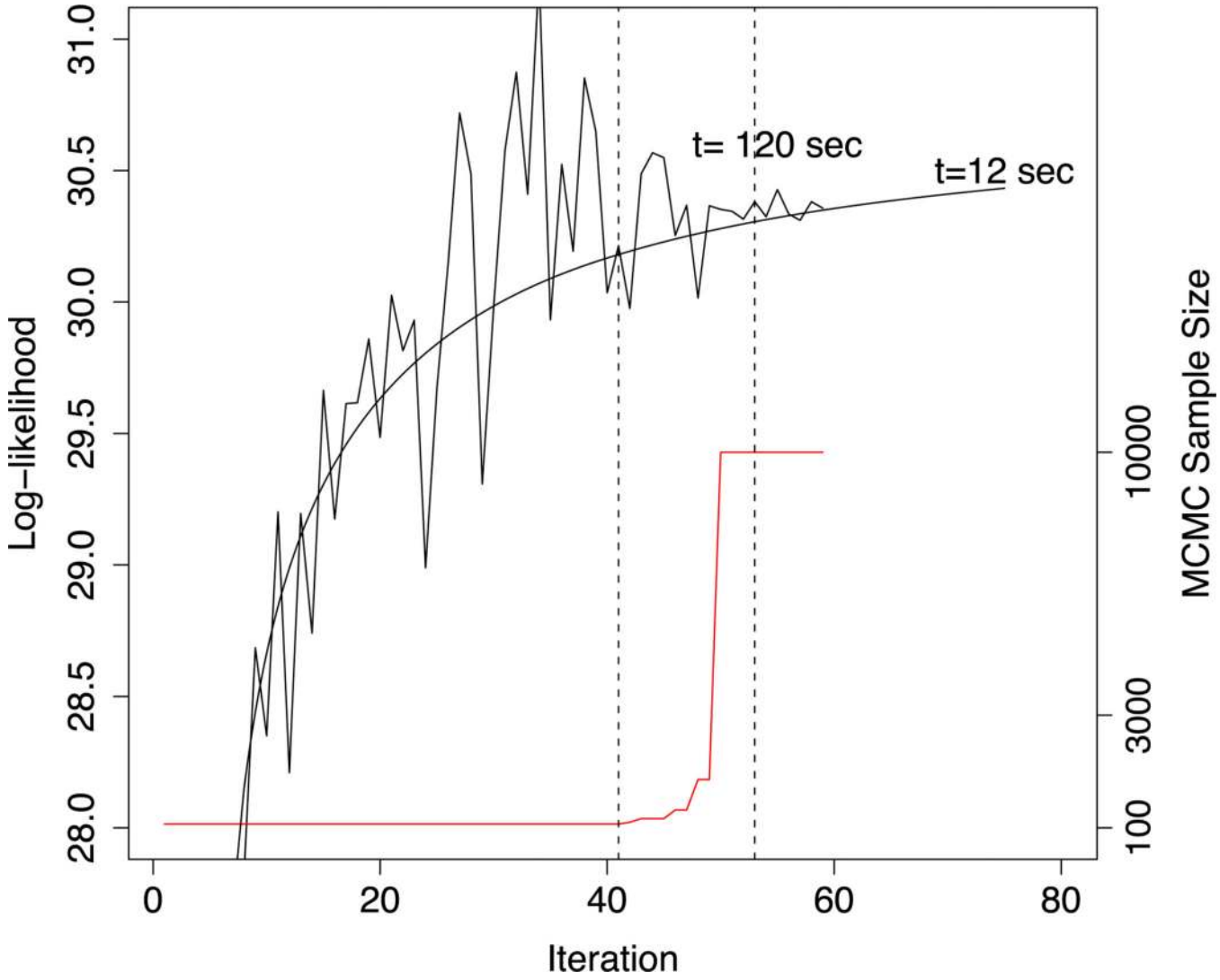

Figure 5.

Comparison of the convergence of MCEM (jagged line) and the proposed EM (smooth line). The convergence times were 120 and 12 sec in this example. The MCMC sample size for the MCEM is plotted at the bottom of the graph. The y-axis represents the surrogate log-likelihood (objective function) given by formula (2.13).

Table 1

Analysis of primary HIV infection. The parameters are for the random intercept logistic model and logistic model with random intercept and random linear decrease after 50 days, respectively.

	Setpoint model		Five-parameter model	
	Estimate	**SE**	**Estimate**	**SE**
β_1	1.575	0.014	1.609	0.014
β_2	0.4240	0.0933	0.1441	0.0950
β_3	3.561	0.034	3.526	0.024
β_4	1.547	0.228	1.060	0.267
β_5			$-3.48 \cdot 10^{-3}$	$1.43 \cdot 10^{-3}$
σ	0.554		0.512	
σ_{b1}	0.139		0.133	
σ_{b5}			$7.10 \cdot 10^{-3}$	
ρ_{b12}			0.17	

Table 2

Linear case: Estimated parameters from a simulation study of 4000 datasets (mean, standard deviation, numeric error), and CPU time in seconds (mean, standard deviation). Three algorithms are compared: the proposed Closed-form EM; Monte Carlo EM; and the LME for complete data. Maximum likelihood estimation is used throughout.

		Closed-form EM			MCEM			Complete data LME	
	True value	**Mean**	**SD**	**NErr**	**Mean**	**SD**	**NErr**	**Mean**	**SD**
Time		5.51	2.95		216	18		0.11	0.002
β_0	−2.83	−2.8297	0.0394	$<10^{-5}$	−2.8298	0.0393	0.0006	−2.8300	0.0344
β_1	−0.18	−0.1800	0.0071	<0.0001	−0.1800	0.0071	0.0019	−0.1799	0.0062
σ	0.15	0.1493	0.0088	0.0017	0.1495	0.0088	0.0024	0.1496	0.0074
$\Psi_{11} \times 10^2$	4.9	4.775	1.540	0.0081	4.743	1.544	0.0153	4.778	1.186
$\Psi_{12} \times 10^3$	1	0.640	1.926	0.0557	0.685	1.935	0.0981	0.616	1.489
$\Psi_{22} \times 10^3$	2	1.773	0.498	0.0041	1.766	0.499	0.0165	1.775	0.377

Table 3

Linear case: Coverage probability and relative bias (%) based on a simulation study of 4000 datasets. The coverage probability has a standard error of 0.004; $z(\rho_{12})$ is the z-transformed correlation coefficient in Ψ.

	Closed-form EM	MCEM	Complete data LME
β_0 (coverage probability)	0.936	0.933	0.943
β_1 (coverage probability)	0.938	0.938	0.941
β_0	0.01	0.01	0.00
β_1	0.02	0.02	0.06
σ	−0.45	−0.31	−0.24
Ψ_{11}	−2.54	−3.20	−2.48
Ψ_{12}	−78.67	−77.16	−79.47
Ψ_{22}	−11.33	−11.68	−11.23
$z(\rho_{12})$	−26.87	−26.13	−28.88

Table 4

Nonlinear case: Estimated parameters from a simulation study of 1000 datasets (mean, standard deviation), and CPU time in seconds (mean, standard deviation), using maximum likelihood.

		Closed-form EM		MCEM		Complete data NLME	
	True value	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
CPU time (sec)		67	12	403	57	3.6	5.7
β_1	1.6094	1.6089	0.0117	1.60840	0.0116	1.6131	0.0110
β_2	0.6931	0.7064	0.0482	0.7091	0.0477	0.6854	0.0423
β_3	3.8067	3.7997	0.0246	3.7996	0.0246	3.7998	0.0238
β_4	2.3026	2.3366	0.1166	2.3434	0.1151	2.2872	0.1078
σ	0.5500	0.5455	0.0147	0.5453	0.0147	0.5485	0.0144
Ψ_{11}	0.0025	0.0029	0.0006	0.0029	0.0006	0.0028	0.0006
Ψ_{12}	−0.0010	−0.0018	0.0041	−0.0018	0.0040	−0.0014	0.0035
Ψ_{22}	0.0100	0.0493	0.0276	0.0532	0.0263	0.0178	0.0285

J Comput Graph Stat. Author manuscript; available in PMC 2015 March 29.

Table 5

Nonlinear case: Coverage probability and relative bias (%) based on a simulation study of 1000 datasets. The coverage probability has a standard error of 0.008.

	Closed-form EM	MCEM	Complete data NLME
β_1 (coverage probability)	0.945	0.944	0.931
β_2 (coverage probability)	0.920	0.921	0.911
β_3 (coverage probability)	0.947	0.949	0.935
β_4 (coverage probability)	0.922	0.925	0.714
β_1	−0.03	−0.06	0.23
β_2	1.91	2.30	−1.12
β_3	−0.18	−0.19	−0.18
β_4	1.48	1.77	−0.67
σ	−0.81	−0.86	−0.27
Ψ_{11}	15.96	16.19	13.10
Ψ_{12}	−83.69	−84.79	−37.66
Ψ_{22}	392.61	431.64	79.91