ELSEVIER Computational Statistics & Data Analysis 25 (1997) 465–490

COMPUTATIONAL STATISTICS & DATA ANALYSIS

Two Taylor-series approximation methods for nonlinear mixed models

Russell D. Wolfinger [a,*], Xihong Lin [b]

[a] *SAS Institute Inc. R-52, SAS Campus Drive, Cary, NC 27513, USA*
[b] *Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA*

Received 1 November 1995; received in revised form 7 February 1997; accepted 27 February 1997

Abstract

This paper considers a general class of nonlinear mixed models (NLMMs). In view of the often difficult numerical integrations involved in a full likelihood analysis of general NLMMs, we discuss two approximate inference procedures, both based on Taylor-series approximations to the integrated likelihood. The two methods differ in the Taylor-series expansion point for the random-effects parameters: the first method expands around zero (the expected value of the random effects) and the second around their empirical best linear unbiased predictors (EBLUPs). We present new algorithms for implementing both approximations and relate them to existing algorithms. Finally, we assess the performance of both methods through simulation, and conclude that the EBLUP-expansion method offers minor gains in accuracy over the zero-expansion method at the cost of higher computing times and instability. © 1997 Elsevier Science B.V.

Keywords: BLUP; Laplace expansion; Likelihood approximation; Random effects; REML; Variance components

AMS Classification: 62J99

1. Introduction

Nonlinear mixed models (NLMMs) have important uses in various disciplines such as pharmacokinetics and growth studies in agriculture and medicine (Sheiner and Beal, 1980; Vonesh and Carter, 1992). These models entertain unknown mean-zero random variables known as random effects, conditional upon which the observed data are assumed to have a Gaussian distribution with some general nonlinear mean

* Corresponding author.

PII S0167-9473(97)00012-1

Table 1
Comparison of the zero-expansion method with the literature

Sheiner and Beal (1980, first-order method): They maximize an objective function based on a zero-expansion, whereas we iteratively solve a set of GEEs.
Longford (1988), Goldstein (1991), Breslow and Clayton (1993 MQL) Wolfinger and O'Connell (1993, population-averaged method): Similar algorithms to ours, but developed in the context of GLMMs. They use a different scaling for the working data vector and weights based on the link function.
Hirst et al. (1991): Very similar to our algorithm, except they use an EM algorithm for the linear mixed model fits (we use Newton–Raphson) and they use nonlinear least squares between fits whereas we solve the mixed model equations. Their algorithm is effectively the same as the approximate maximum likelihood technique of Lindstrom and Bates' except the expansion is about zero instead of the EBLUPs.
Vonesh and Carter (1992): Similar iterative scheme, except they use method of moments at each stage to estimate the variance covariance parameters whereas we maximize a working likelihood or residual likelihood.
Solomon and Cox (1992, Section 4): They consider higher-order Taylor series expansions about zero whereas we consider only a second-order expansion. They do not focus on estimation, but they appear to have in mind a direct maximization of the approximate likelihood as in Sheiner and Beal's first-order method.

function and unknown variance–covariance parameters. They are closely related to generalized linear mixed models (GLMMs) (refer to Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993) which extend familiar exponential family models to include random effects. While GLMMs accommodate non-normally distributed data, they typically also postulate a more restrictive nonlinear conditional mean structure than the general one we consider under NLMMs.

Unlike their linear counterpart, a full likelihood analysis of NLMMs often requires difficult numerical integrations. Although some success has been documented with Gauss–Hermite quadrature (Davidian and Gallant, 1993) and Markov Chain Monte Carlo (Zeger and Karim, 1991; Wakefield, 1996), straightforward linearization techniques appear to offer the analyst basically reliable results in many contexts. We develop two such techniques herein.

The two techniques are both based on a Taylor-series expansion of the likelihood around values for the random-effects parameters. The first technique, called the zero-expansion method, expands around the zero vector, which is the assumed expected value of the random effects. The second technique, called the EBLUP-expansion method, expands around the empirical best linear unbiased predictors of the random effects (refer to Harville, 1990; Robinson, 1991). Both approximations lead to algorithms involving the iterative fitting of linear mixed models to suitably constructed pseudo-data. The linear mixed model fits can be performed using either standard maximum likelihood (ML) or residual maximum likelihood (REML). Each procedure can be viewed as iteratively solving a certain set of generalized estimating equations (GEEs), a concept popularized in the generalized linear model setting by Liang and Zeger (1986), Zeger and Liang (1986), and Zeger et al. (1988).

Both of the aforementioned expansions have already been proposed in the literature, although not in the general and unified fashion we consider here. To place our

Table 2
Comparison of the EBLUP-expansion method with the literature

Lindstrom and Bates (1990), Pinheiro and Bates (1995, LME Approximation): We solve the same set of GEEs, but our algorithms differ. They alternate between linear mixed model and nonlinear least-squares steps, whereas we fit successive linear mixed models and the accompanying mixed model equations. Also, they assume an additive form for the fixed and random effects, whereas we have a general nonlinear form.
Schall (1991), Breslow and Clayton (1993 PQL), Wolfinger and O'Connell (1993, subject-specific method), Engel and Keen (1994), McGilchrist (1994), Tempelman and Gianola (1995, MML): Similar algorithms to ours, but developed in the context of GLMMs, and each article uses a different analytical motivation. Also, the algorithms employ iteratively updated weights based on the GLMM link function as well as different scalings for the working data.
Beal and Sheiner (1992, first-order conditional method): They maximize and objective function based on an EBLUP-expansion whereas we iteratively solve a set of GEEs.
Vonesh (1996): GEEs are different; we set certain off-diagonal matrices in his GEEs to zero. Vonesh's proposed estimates appear to be the same as those produced by Beal and Sheiner's first-order conditional method.

developments in context, Tables 1 and 2 provide a comparison of our methods with many existing techniques. Descriptions of many of the terms used in these tables are found in later sections. Table 1 covers the zero-expansion method, and the closest method in the literature is that of Hirst et al. (1991). However, they consider only a ML approximation whereas we develop both ML and REML approaches. Also, they use an EM algorithm to maximize the linear mixed model likelihood at each iteration whereas we use Newton–Raphson. Even so, using our zero-expansion method with ML should produce the same results as the Hirst et al. (1991) modulo convergence tolerance differences.

Table 2 lists references relating to our EBLUP-expansion method. The closest technique in the literature here is that of Lindstrom and Bates (1990), although the list of references beginning with Schall (1991) provide very similar analyses for GLMMs. Our EBLUP-expansion method should produce the same estimates as the Lindstrom and Bates method even though the algorithms differ.

More detailed comparisons with the literature are found in Sections 2–4, in which we present an analytical development of the zero- and EBLUP-expansion methods. Algorithms for each are motivated using Laplace approximations, a particular application of Taylor-series expansions well suited to the likelihood framework we consider.

Since both procedures are approximate likelihood methods, they can be expected to perform well only under conditions for which the approximations provide reasonable accuracy. Since such conditions can be difficult to explicitly identify in practice, simulation studies are integral in assessing the general performance of such techniques in practice. Sections 5 and 6 present such simulation studies for two common NLMMs: a logistic growth curve and a one-compartment model from pharmacokinetics. Section 7 provides additional discussion of the results.

2. The nonlinear mixed model

Consider an observation vector y of length n together with matrices $X_{n\times r}$ and $Z_{n\times s}$ of explanatory variables associated with the fixed and random effects, respectively. The general NLMM is defined as

$$y = f(X, \beta, Z, b) + e, \tag{1}$$

where β is a $p \times 1$ vector of unknown fixed-effects parameters, b is a $q \times 1$ vector of unknown random-effects variables distributed as $\mathrm{N}(0, D)$, and e is an unknown vector of residual disturbances following $\mathrm{N}(0, R)$ independently of b. The elements of D and R are assumed to be functions of an unknown vector θ of variance and covariance parameters.

This class of models includes various models considered in the literature as special cases. Vonesh and Carter (1992) and Gumpertz and Pantula (1992) consider cases where random effects enter the model additively:

$$y = f(X, \beta) + Zb + e. \tag{2}$$

When f is invertible, the model

$$y = f(X\beta + Zb) + e$$

can be regarded as a special case of a GLMM applied to normally distributed data with nonidentity link function. Lindstrom and Bates (1990) and Wolfinger (1993) develop notation in terms of the additive form $X\beta + Zb$, although with appropriate construction this form is not necessarily very restrictive. Note that in the repeated measures setting considered by these authors, $D(\theta)$ and $R(\theta)$ are block diagonal matrices with the ith block corresponding to the ith cluster.

We develop likelihood-based methods in terms of the general specification (1) to explicitly handle models that are expressible only in this form. Examples include crossed and nested designs as well as the one-compartment model considered in Section 6. The integrated likelihood of (1) for (β, θ) is

$$L(\beta, \theta) = \mathrm{e}^{\ell(\beta,\theta)} = (2\pi)^{-q/2}|D|^{-1/2} \int \mathrm{e}^{\ell(y;b) - (1/2)b^{\mathrm{T}}D^{-1}b}\,\mathrm{d}b, \tag{3}$$

where

$$\ell(y; b) = -\tfrac{1}{2}n\ln 2\pi - \tfrac{1}{2}\ln|R| - \tfrac{1}{2}(y - f(X, \beta, Z, b))^{\mathrm{T}}R^{-1}(y - f(X, \beta, Z, b)).$$

Except for model (2), a closed-form expression of $L(\beta, \theta)$ is typically not available, and a full likelihood analysis is hampered by the need for numerical integration. We therefore investigate two Taylor-series approximation methods. The first, described in the following section, uses $E(b) = 0$ as the expansion point. The second, described in Section 4, uses successive EBLUPs of b as expansion points.

3. The zero-expansion method

In the tradition of the pioneering work by Sheiner and Beal (1980), a natural mathematical approximation for the integrand in (3) is a Taylor-series expansion about $b=0$ before integration. Solomon and Cox (1992) study this approximation for NLMMs with a single small variance component, and we now generalize their results to model (1). A quadratic expansion yields the approximate loglikelihood, which contains the leading terms of the Laplace approximation of Barndorff-Nielsen and Cox (1989, Eq. (3.16)), as follows:

$$\ell_0(\beta,\theta) \approx -\tfrac{1}{2}n\ln 2\pi - \tfrac{1}{2}\ln|V_0| - \tfrac{1}{2}(y - f_0(X,\beta))^{\mathrm{T}} V_0^{-1}(y - f_0(X,\beta)). \tag{4}$$

Here

$$V_0 = Z_0^* D Z_0^{*\mathrm{T}} + R,$$

$$Z_0^* = \partial f/\partial b^{\mathrm{T}}|_{b=0},$$

$$f_0(X,\beta) = f(X,\beta,Z,b)|_{b=0},$$

and we have approximated $\partial^2\ell(y;b)/\partial b\partial b^{\mathrm{T}}|_{b=0}$ by $-Z_0^{*\mathrm{T}}R^{-1}Z_0^*$ and applied the identities

$$V_0^{-1} = R^{-1} - R^{-1}Z_0^* D(I + Z_0^{*\mathrm{T}}R^{-1}Z_0^* D)^{-1}Z_0^{*\mathrm{T}}R^{-1}, \tag{5}$$

$$|V_0| = |R|\left|I + Z_0^{*\mathrm{T}}R^{-1}Z_0^* D\right| \tag{6}$$

(refer to Harville, 1977, Eqs. (3.6) and (5.1)).

This approximation in fact corresponds to the loglikelihood of y in the mixed model

$$y = f_0(X,\beta) + Z_0^* b + e, \tag{7}$$

where $b \sim \mathrm{N}(0,D)$ and $e \sim \mathrm{N}(0,R)$ and is of the same form as model (2). The mixed model (7) can also be derived by taking a one-term Taylor expansion of the conditional mean $f(X,\beta,Z,b)$ in (1) about $b=0$. Gumpertz and Pantula (1992) provide a thorough treatment of the special case of (7) where Z_0^* does not depend upon β, a case where exact likelihood inference is analytically tractable.

Sheiner and Beal (1980) numerically optimize (4) over β and θ in their well-known first-order method. Instead, we use (4) to derive the iterative zero-expansion estimation procedure for β, b, and θ. Our strategy is to first fix θ to estimate β, to next use the estimate for β to derive an updated estimate for θ, and then to alternate in this fashion until convergence. Zero-expansion predictions for b can be obtained from the final estimates of β and θ using the standard EBLUP formula described in the next subsection.

3.1. Estimation of the fixed- and random-effects parameters

We fix θ for the moment and ignore the dependence of V_0 on β through Z_0^* in differentiating $\ell_0(\beta,\theta)$ in (4) with respect to β. This approximation can be justified by an appeal to intrinsic as opposed to parameter-effects nonlinearity (Bates and Watts, 1980). The arguments of these authors show that the space spanned by the columns of Z_0^* depends only on the intrinsic curvature of the nonlinear model, but not on the parameter-effects curvature in the tangent plane. Therefore, Z_0^* may be assumed to vary somewhat slowly with β. We thus estimate β by $\hat\beta_0$, which solves

$$X_0^{*\mathrm{T}} V_0^{-1}(y - f_0(X,\hat\beta_0)) = 0, \tag{8}$$

where

$$X_0^* = \partial f_0/\partial\beta. \tag{9}$$

Equivalently, $\hat\beta_0$ solves

$$\hat\beta_0 = (X_0^{*\mathrm{T}} V_0^{-1} X_0^*)^{-1} X_0^{*\mathrm{T}} V_0^{-1} Y_0, \tag{10}$$

where $Y_0 = y - f_0(X,\hat\beta_0) + X_0^*\hat\beta_0$, and thus $\hat\beta_0$ is a standard generalized least-squares estimate using pseudo-data. In like fashion, the covariance matrix of $\hat\beta_0$ can be approximated as

$$\mathrm{cov}(\hat\beta) \approx (X_0^{*\mathrm{T}} V_0^{-1} X_0^*)^{-1}. \tag{11}$$

In the zero-expansion method the prediction of b can be deferred until final estimates for β and θ have been obtained. Assuming $\hat\beta_0$ and θ are set equal to their final estimates, b is predicted by the EBLUP formula

$$\hat b_0 = D Z_0^{*\mathrm{T}} V_0^{-1}(Y_0 - X_0^*\hat\beta_0) \tag{12}$$

with prediction covariance estimated by

$$\mathrm{cov}(\hat b_0 - b) \approx D - D Z_0^{*\mathrm{T}} V_0^{-1} Z_0^* D + D Z_0^{*\mathrm{T}} V_0^{-1} X_0^* (X_0^{*\mathrm{T}} V_0^{-1} X_0^*)^{-1} X_0^{*\mathrm{T}} V_0^{-1} Z_0^* D. \tag{13}$$

Expressions (11) and (13) follow linear mixed model theory (see, for example, Laird and Ware, 1982). However, analysts should be aware that these estimators, including analogous ones for the EBLUP-expansion method, ignore the uncertainty in estimating θ. The common circumvention of this difficulty in linear mixed models involves the use of t- and F-statistics with approximate degrees of freedom, possibly coupled with some sort of bias adjustment (e.g. Kackar and Harville, 1984). However, standard approximations such as Satterthwaite's method and adjustments based on asymptotic orthogonality could be unreliable in the nonlinear case. Simulation studies appear necessary to evaluate the severity of these potential inaccuracies in general, and this is one of the issues we address in our simulation studies later in the paper.

3.2. Estimation of the variance parameters

The ML zero-expansion estimation of θ proceeds by maximizing the approximate profile loglikelihood of θ constructed by substituting $\beta = \hat{\beta}_0(\theta)$ into (4). Specifically,

$$\ell_0^*(\theta) = -\tfrac{1}{2}n\ln 2\pi - \tfrac{1}{2}\ln|V_0| - \tfrac{1}{2}(Y_0 - X_0^*\beta)^{\mathrm{T}}V_0^{-1}(Y_0 - X_0^*\beta)\Big|_{\beta=\hat{\beta}_0(\theta)}. \tag{14}$$

The derivative of (14) with respect to θ yields the ML zero-expansion estimating equations for θ:

$$-\frac{1}{2}\mathrm{tr}\left(V_0^{-1}\frac{\partial V_0}{\partial\theta_j}\right) + \frac{1}{2}\left(Y_0 - X_0^*\hat{\beta}_0\right)^{\mathrm{T}}V_0^{-1}\frac{\partial V_0}{\partial\theta_j}V_0^{-1}(Y_0 - X_0^*\hat{\beta}_0) = 0. \tag{15}$$

Note, the dependence of Z_0^* on θ is ignored in calculating $\partial V_0/\partial\theta_j$ so that $\partial V_0/\partial\theta_j = Z^*\partial D/\partial\theta_j Z^{*\mathrm{T}} + \partial R/\partial\theta_j$. The jkth element of the associated information matrix is $\frac{1}{2}\mathrm{tr}(V_0^{-1}\,\partial V_0/\partial\theta_j V_0^{-1}\partial V_0/\partial\theta_k)$.

Solving (10) and (15) iteratively until convergence yields the ML zero-expansion estimates, which may be viewed as approximate ML estimates of (β,θ). Simultaneous solutions to (10) and (15) for each iteration are achieved numerically via standard linear mixed model routines using the working vector Y_0 and the working design matrices X_0^* and Z_0^* in the form $Y_0 = X_0^*\beta + Z_0^*b + e$, where $b \sim \mathrm{N}(0, D(\theta))$ and $e \sim \mathrm{N}(0, R(\theta))$. Note that both (10) and (15) can be viewed as GEEs, establishing a connection with GLMM methods such as the population-averaged method of Wolfinger and O'Connell (1993).

In order to take into account the loss of degrees of freedom from estimating β, the residual profile loglikelihood and its corresponding estimating equations for θ are often preferred in practice (Harville, 1977). Specifically, the REML version of (14) is

$$\begin{aligned}\ell_{0R}^*(\theta) = &-\tfrac{1}{2}n\ln 2\pi - \tfrac{1}{2}\ln|X_0^{*\mathrm{T}}V_0^{-1}X_0^*| \\ &- \tfrac{1}{2}\ln|V_0| - \tfrac{1}{2}(Y_0 - X_0^*\beta)^{-1}V_0^{-1}(Y_0 - X_0^*\beta)\Big|_{\beta=\hat{\beta}(\theta)}.\end{aligned} \tag{16}$$

Assuming that the derivative matrix X_0^* varies slowly with (β,θ) (Bates and Watts, 1980), the derivative of (16) with respect to θ leads to the REML zero-expansion estimating equations for θ:

$$-\frac{1}{2}\mathrm{tr}\left(P_0\frac{\partial V_0}{\partial\theta_j}\right) + \frac{1}{2}(Y_0 - X_0^*\hat{\beta})^{\mathrm{T}}V_0^{-1}\frac{\partial V_0}{\partial\theta_j}V_0^{-1}(Y_0 - X_0^*\hat{\beta}) = 0. \tag{17}$$

where

$$P_0 = V_0^{-1} - V_0^{-1}X_0^*(X_0^{*\mathrm{T}}V_0^{-1}X_0^*)^{-1}X_0^{*\mathrm{T}}V_0^{-1}. \tag{18}$$

The components of the corresponding information matrix are

$$I(j,k) = \frac{1}{2}\,\mathrm{tr}\left(P_0\frac{\partial V_0}{\partial\theta_j}P_0\frac{\partial V_0}{\partial\theta_k}\right). \tag{19}$$

The REML zero-expansion method alternates between solving (10) and (17). Expression (17) can be solved using REML estimation in standard linear mixed models software, and (10) is simply estimated generalized least-squares based on the current REML estimate of θ.

4. The EBLUP-expansion method

We now proceed to the second of the estimation procedures considered in this article, the EBLUP-expansion method. We base it on Laplace's approximation, in which an integral of the form $\int e^{-g(b)}\,db$ is approximated by making a quadratic Taylor-series expansion of $-g(b)$ about its maximum value $\tilde{b}$ before integration (refer to Tierney and Kadane, 1986; Barndorff-Nielsen and Cox, 1989; Breslow and Clayton, 1993; Wolfinger, 1993). It follows that the loglikelihood $\ell(\beta,\theta)$ in (3) is approximated by

$$\ell_E(\beta,\theta) = -\tfrac{1}{2}n\ln 2\pi - \tfrac{1}{2}(\ln|\tilde{Z}^{*\mathrm{T}}R^{-1}\tilde{Z}^{*}D + I| + \ln|R|)$$
$$- \tfrac{1}{2}(y - f(X,\beta,Z,\tilde{b}))^{\mathrm{T}}R^{-1}(y - f(X,\beta,Z,\tilde{b})) - \tfrac{1}{2}\tilde{b}^{\mathrm{T}}D^{-1}\tilde{b}, \tag{20}$$

where $Z^* = \partial f/\partial b^{\mathrm{T}}$, $\tilde{Z}^* = Z^*|_{b=\tilde{b}}$ and $\tilde{b}$ solves

$$\frac{\partial g}{\partial b} = -Z^{*\mathrm{T}}R^{-1}(y - f(X,\beta,Z,b)) + D^{-1}b = 0. \tag{21}$$

We here have used the approximation

$$\frac{\partial^2 g}{\partial b\partial b^{\mathrm{T}}} = Z^{*\mathrm{T}}R^{-1}Z^* + D^{-1} - \sum_{i=1}^{n}\frac{\partial^2 f_i}{\partial b\partial b^{\mathrm{T}}}r_i$$
$$\approx Z^{*\mathrm{T}}R^{-1}Z^* + D^{-1},$$

where r_i is the ith component of the $n\times 1$ residual vector $R^{-1}(y - f(X,\beta,Z,b))$, which has mean zero. This Laplace approximation is the same as the one discussed by Vonesh (1996) but is different from that given in Wolfinger (1993), where the expansion was taken about the mode of both β and b under a flat prior of β.

Denoting $V = Z^*DZ^{*\mathrm{T}} + R$ and $\tilde{V} = V|_{b=\tilde{b}}$, and applying analogs to (5) and (6), $\ell_E(\beta,\theta)$ can be expressed as follows:

$$\ell_E(\beta,\theta) = -\tfrac{1}{2}n\ln 2\pi - \tfrac{1}{2}\ln|\tilde{V}| - \tfrac{1}{2}(y - f(X,\beta,Z,\tilde{b}))^{\mathrm{T}}$$
$$\times R^{-1}(y - f(X,\beta,Z,\tilde{b})) - \tfrac{1}{2}\tilde{b}^{\mathrm{T}}D^{-1}\tilde{b} \tag{22}$$
$$= -\tfrac{1}{2}n\ln 2\pi - \tfrac{1}{2}\ln|\tilde{V}| - \tfrac{1}{2}(y - f(X,\beta,Z,\tilde{b}) + \tilde{Z}^*\tilde{b})^{\mathrm{T}}$$
$$\times \tilde{V}^{-1}(y - f(X,\beta,Z,\tilde{b}) + \tilde{Z}^*\tilde{b}). \tag{23}$$

This approximation is exact for model (2) in which the random effects enter linearly. Lindstrom and Bates (1990, Eq. (4.3)) derived the same approximation by linearizing $f(X,\beta,Z,b)$ in (1) about $b = \tilde{b}$.

Expression (23) is now used as a basis for deriving the EBLUP-expansion estimation procedure. The acronym EBLUP is used in association with this scheme because of the fact that $\tilde{b}$ in each iteration is actually the empirical best linear unbiased predictor of the random-effects parameter vector b utilized in the working linear mixed model described in the following subsections.

4.1. Estimation of the fixed- and random-effects parameters

The development of the EBLUP-expansion method is in many ways similar to the foregoing derivation of the zero-expansion method, and the same basic strategy is adopted here. We begin with a similar appeal to intrinsic nonlinearity (Bates and Watts, 1980), in order to ignore the dependence of $\tilde{V}$ on β through $\tilde{Z}^*$. Thus, the value of β that maximizes (22) for fixed θ can be equivalently obtained by maximizing a log penalized quasilikelihood (Green, 1987; Breslow and Clayton, 1993) with respect to (β, b):

$$-\tfrac{1}{2}(y - f(X,\beta,Z,b))^{\mathrm{T}}R^{-1}(y - f(X,\beta,Z,b)) - \tfrac{1}{2}b^{\mathrm{T}}D^{-1}b. \tag{24}$$

Differentiating (24) with respect to (β, b) yields the estimating equations

$$\begin{aligned} X^{*\mathrm{T}}R^{-1}(y - f(X,\beta,Z,b)) &= 0, \\ Z^{*\mathrm{T}}R^{-1}(y - f(X,\beta,Z,b)) &= D^{-1}b, \end{aligned} \tag{25}$$

where $X^* = \partial f/\partial\beta^{\mathrm{T}}$. Denoting $Y = y - f(X,\beta,Z,b) + X^*\beta + Z^*b$ and applying the Fisher scoring algorithm to (25) lead to iteratively solving the following equations:

$$\begin{bmatrix} X^{*\mathrm{T}}R^{-1}X^* & X^{*\mathrm{T}}R^{-1}Z^* \\ Z^{*\mathrm{T}}R^{-1}X^* & Z^{*\mathrm{T}}R^{-1}Z^* + D^{-1} \end{bmatrix} \begin{pmatrix} \beta \\ b \end{pmatrix} = \begin{bmatrix} X^{*\mathrm{T}}R^{-1}Y \\ Z^{*\mathrm{T}}R^{-1}Y \end{bmatrix}. \tag{26}$$

These are the well-known mixed model equations (Henderson, 1984) and are equivalent to estimating β and b by the familiar formulas

$$\begin{aligned} \hat{\beta} &= (X^{*\mathrm{T}}V^{-1}X^*)^{-1}X^{*\mathrm{T}}V^{-1}Y, \\ \hat{b} &= DZ^{*\mathrm{T}}V^{-1}(Y - X^*\hat{\beta}). \end{aligned}$$

Wolfinger (1993) shows how these estimates are also one-step Gauss–Newton estimators in the nonlinear least-squares step of Lindstrom and Bates (1990). The approximate variance–covariance matrix of $(\hat{\beta}, \hat{b})$ is the generalized inverse of the coefficient matrix in (26), the diagonal terms of which are as follows:

$$\begin{aligned} \mathrm{cov}(\hat{\beta}) &\approx (X^{*\mathrm{T}}V^{-1}X^*)^{-1} \\ \mathrm{cov}(\hat{b} - b) &\approx D - DZ^{*\mathrm{T}}V^{-1}Z^*D + DZ^{*\mathrm{T}}V^{-1}X^*(X^*V^{-1}X^*)^{-1}X^{*\mathrm{T}}V^{-1}Z^*D. \end{aligned} \tag{27}$$

4.2. Estimation of the variance parameters

The ML EBLUP-expansion estimation of θ proceeds by maximizing the approximate profile loglikelihood of θ constructed by substituting $\beta = \hat{\beta}(\theta)$ into (23). Specifically,

$$\ell_E^*(\theta) = -\tfrac{1}{2}n\ln 2\pi - \tfrac{1}{2}\ln|V| - \tfrac{1}{2}(Y - X^*\beta)^{\mathrm{T}}V^{-1}(Y - X^*\beta)\Big|_{\beta=\hat{\beta}(\theta)}, \tag{28}$$

the EBLUP-expansion analog of (14). The derivative of (28) with respect to θ yields the ML EBLUP-expansion estimating equations for θ:

$$-\frac{1}{2}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial\theta_j}\right) + \frac{1}{2}(Y - X^*\hat{\beta})^{\mathrm{T}}V^{-1}\frac{\partial V}{\partial\theta_j}V^{-1}(Y - X^*\hat{\beta}) = 0. \tag{29}$$

Note the dependence of Z^* on θ is ignored in calculating $\partial V/\partial\theta_j$ so that $\partial V/\partial\theta_j = Z^*\partial D/\partial\theta_j Z^{*\mathrm{T}} + \partial R/\partial\theta_j$. The jkth element of the associated information matrix is $\frac{1}{2}\,\mathrm{tr}(V^{-1}\,\partial V/\partial\theta_j V^{-1}\partial V/\partial\theta_k)$.

Solving (25) and (29) iteratively yields the ML EBLUP-expansion estimates. As with the ML zero-expansion method, this can be accomplished by fitting a linear mixed model to the working response vector Y and the working design matrices X^* and Z^* in the form $Y = X^*\beta + Z^*b + e$, where $b \sim \mathrm{N}(0, D(\theta))$ and $e \sim \mathrm{N}(0, R(\theta))$.

The REML EBLUP-expansion method is based on expressions (17)–(19) but with V_0, Y_0 and X_0^* replaced by V, Y and X^*, respectively. Again, linear mixed models routines can be used to obtain REML-based solutions for θ at each iteration. Given the current estimate of θ, the mixed-model equations (26) are then solved to obtain updated estimates of β and b, and the process is alternated until convergence.

5. Logistic model simulation

In this section and the next we compare through simulation the performance of the zero- and EBLUP-expansion methods discussed in Sections 3 and 4. Both sections add to the simulation results recently given by Pinheiro and Bates (1996). All simulations were conducted on an HP 9000/720 workstation running HP-UX 9.1 with 32 mb of memory using the SAS NLINMIX macro, which iteratively calls the SAS MIXED procedure (refer to SAS Institute Inc., 1996, Littell et al., 1996). The macro is available at http://www.sas.com/techsup/download/stat/ or from SAS Technical Support. REML is used for all PROC MIXED calls, and the algorithms are considered to have converged when the relative differences of all fixed-effects and variance–covariance parameters is less than 1×10^{-8}.

5.1. Model

We expand here upon the logistic model simulation study conducted by Pinheiro and Bates (1995). For the jth observation of the ith subject ($i = 1, \ldots, 15$; $j =$

$1, \ldots, 10$), the model is assumed to be

$$y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp\{-[t_{ij} - (\beta_2 + b_{i2})]/\beta_3\}} + e_{ij},$$

where $t_{ij} = 100, 267, 433, 600, 767, 933, 1100, 1267, 1433, 1600$ for all i. The random effects $b_i = (b_{1i}, b_{2i})^{\mathrm{T}}$ are $\mathrm{N}(0, D)$, and the e_{ij} are $\mathrm{N}(0, \sigma^2)$ and are independent of the b_i. We set $\beta = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}} = (200, 700, 350)^{\mathrm{T}}$,

$$D = \begin{bmatrix} D_{11} & D_{21} \\ D_{21} & D_{22} \end{bmatrix} = \begin{bmatrix} 100 & -50 \\ -50 & 625 \end{bmatrix},$$

and $\sigma^2 = 25$. These values correspond to those used by Pinheiro and Bates (1995).

Both expansion methods are applied to each of 1000 simulated data sets. Additional simulations are created by using the same values of (β, σ^2) and dividing and multiplying each element of D by 25. Therefore, six different simulation runs are performed, resulting from the two estimation methods crossed with the three settings of the variance–covariance parameters. These are denoted small D-Zero, small D-EBLUP, medium D-Zero, medium D-EBLUP, large D-Zero, and large D-EBLUP in the figures and tables that follow. The adjectives small, medium, and large are with reference to the values assumed for D.

The same random number seed is used for all six simulations for comparability purposes. We repeated the entire experiment with a different seed to check for possible spurious effects, but none were found. For both expansion methods (zero, EBLUP), the estimates of D_{11} and D_{22} are constrained to be nonnegative, but the covariance matrix D is not constrained to be nonnegative definite. The final estimates for D are therefore allowed not to be proper variance–covariance matrices. This is done to avoid nonlinear constraints in the linear mixed model phase of the estimation and to keep the estimation parameterized in terms of the original parameters.

Fig. 1 plots example profiles for each of the three sizes of variance components. It illustrates the extent to which the response curves are more variable as the variance components become larger. Also, variability increases with the mean for this particular model.

5.2. *Results for the fixed-effects parameters*

Fig. 2 contains boxplots for the fixed-effects parameter estimates from the six simulation runs. Similar plots are used in Vonesh (1992), with the bottom and top edges of the box located at the 25th and 75th percentiles of the 1000 estimates and the horizontal line in the box at the median. The vertical lines extend from the box as far as the data extend, to a distance of at most 1.5 interquartile ranges. In Fig. 2, the increase in variability with the size of the variance components is clearly evident in the graphs for β_1 and β_2. The plot for β_3 reveals constant variability across the three sizes of variance components and the poorer results for the large D-zero simulation.

Table 3 contains simulation summary statistics for the fixed-effects parameter estimates. In Table 3, True denotes the value used to generate the data, Mean denotes

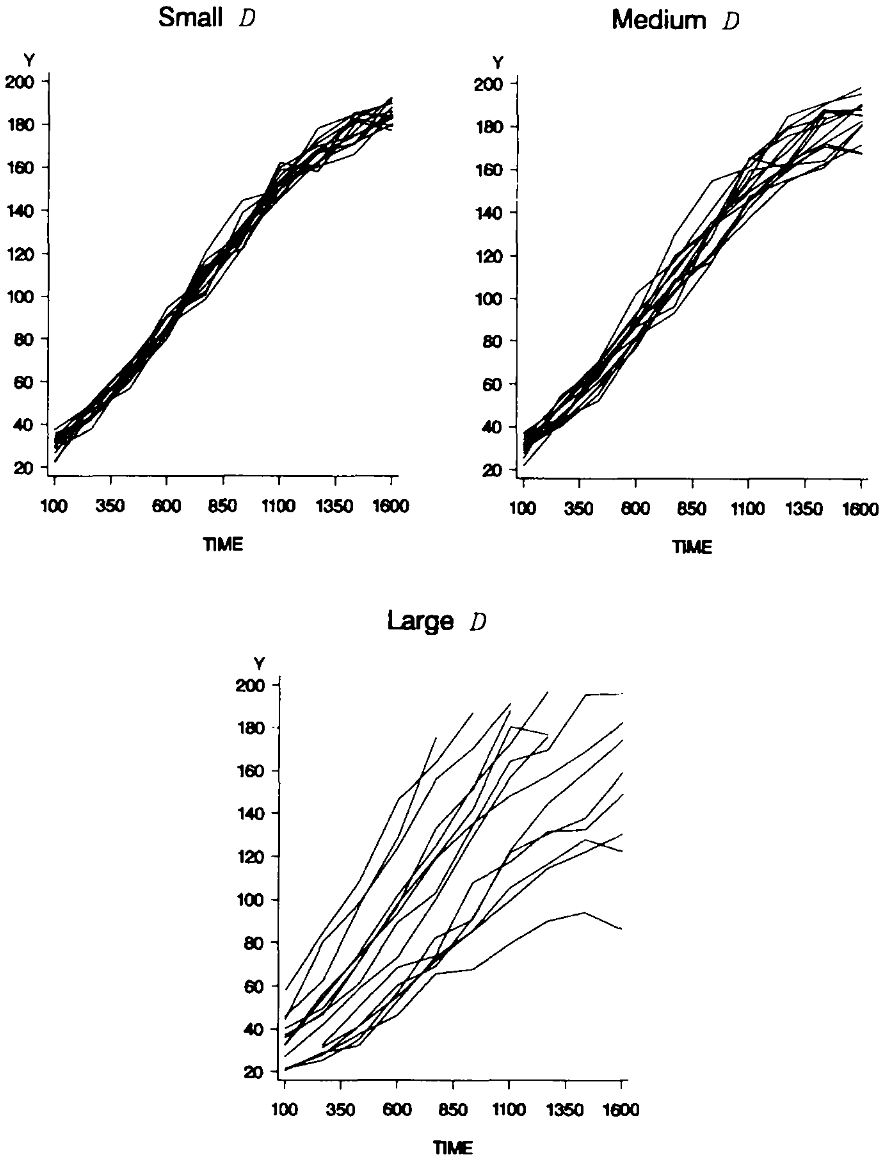

Fig. 1. Example of simulated logistic curves for three sizes of random effects variance (small, medium, large).

the arithmetic average of 1000 estimates, and Rel. Bias equals the relative bias, (Mean − True)/ |True|. Rel. S.D. is the relative standard deviation, which is the standard deviation of the 1000 estimates divided by the absolute value of the true value. It provides a direct measure of the sampling variability of the estimates. Rel. MSD is the relative model standard deviation, which is the average of the model-based standard errors derived from each model fit divided by the absolute value of true value. It can be compared with Rel. S.D. to assess the accuracy of the

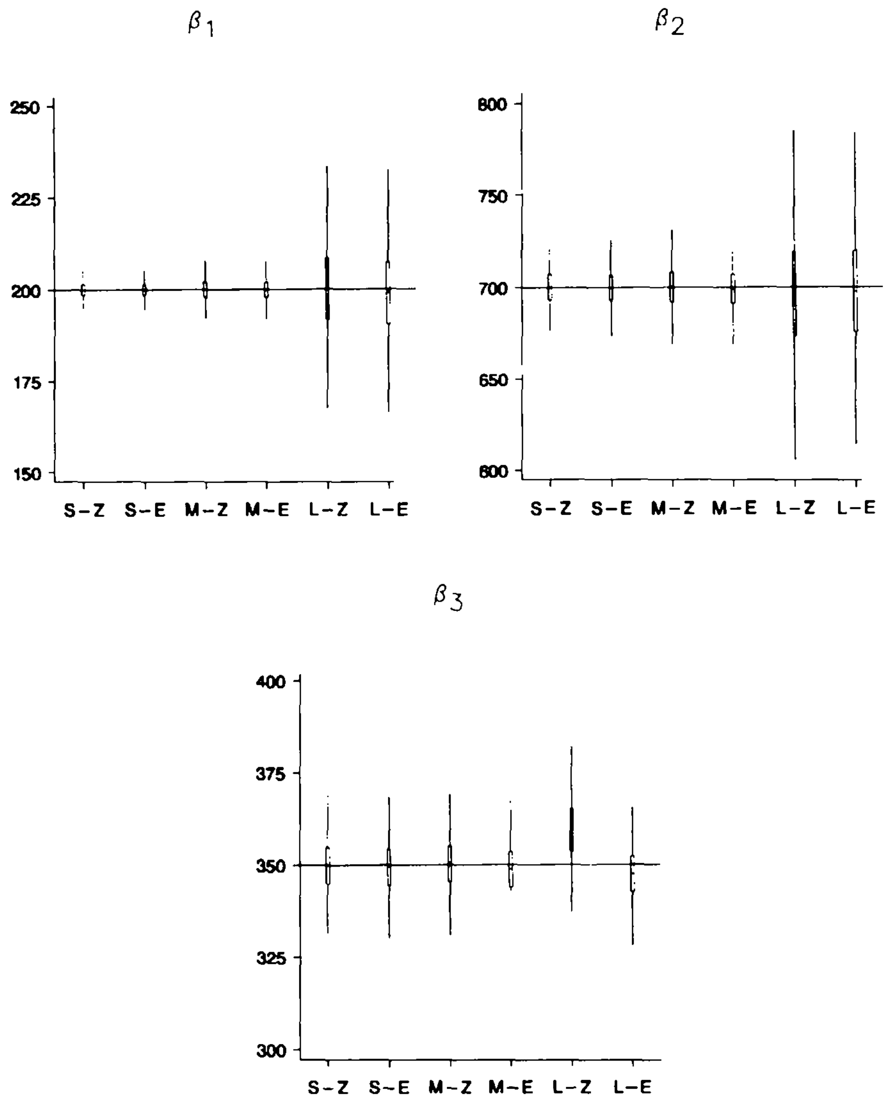

Fig. 2. Box plots of fixed-effects parameter estimates from the simulation settings of the logistic model: S-Z (small D-zero), S-E (small D-EBLUP), M-Z (medium D-zero), M-E (medium D-EBLUP), L-Z (large D-zero), and L-E (large D-EBLUP). The horizontal reference lines are located at the true values of the parameters.

model-based estimates of variability. Rel. RMSE is the relative root mean squared error, and it equals the square root of the mean squared error of the estimate divided by |True|. It provides a standard measure of efficiency. Finally, 95% Cov. denotes the observed coverage of the t-based 95% confidence intervals computed using the model-based standard errors and a t-distribution critical value based on $n - \text{rank}(X||Z) = 119$ degrees of freedom. The 95% coverage values marked with an asterisk in Table 3 are outside the interval (92.93, 97.07). The half-width of this interval is three times the binomial standard error, which is $[(95)(5)/1000]^{1/2} = 0.6892$.

Table 3
Summary statistics for fixed-effects parameter estimates from the simulations of the logistic model

Simulation	True	Mean	Rel. Bias	Rel. S.D.	Rel. MSD	Rel. RMSE	95% Cov.
β_1							
Small *D*-zero	200	200.1	0.0003	0.010	0.010	0.010	94.8
Small *D*-EBLUP	200	199.9	−0.0003	0.010	0.010	0.010	94.8
Medium *D*-zero	200	200.0	0.0000	0.016	0.016	0.016	94.1
Medium *D*-EBLUP	200	199.7	−0.0016	0.016	0.016	0.016	93.9
Large *D*-zero	200	200.1	0.0003	0.063	0.064	0.063	94.1
Large *D*-EBLUP	200	199.1	−0.0047	0.063	0.064	0.063	93.1
β_2							
Small *D*-zero	700	700.5	0.0007	0.015	0.014	0.015	95.2
Small *D*-EBLUP	700	699.9	−0.0002	0.014	0.014	0.014	94.6
Medium *D*-zero	700	700.3	0.0004	0.017	0.017	0.017	94.0
Medium *D*-EBLUP	700	699.0	−0.0014	0.017	0.016	0.017	93.3
Large *D*-zero	700	695.9	−0.0059	0.049	0.049	0.049	93.2
Large *D*-EBLUP	700	697.0	−0.0042	0.047	0.047	0.047	94.1
β_3							
Small *D*-zero	350	350.0	−0.0000	0.022	0.021	0.022	93.6
Small *D*-EBLUP	350	349.5	−0.0014	0.022	0.021	0.022	93.4
Medium *D*-zero	350	350.4	0.0011	0.022	0.021	0.022	93.6
Medium *D*-EBLUP	350	348.8	−0.0034	0.022	0.021	0.022	92.7*
Large *D*-zero	350	359.6	0.0273	0.025	0.022	0.037	78.3*
Large *D*-EBLUP	350	347.4	−0.0073	0.021	0.021	0.023	91.6*

* 95% coverage outside (92.93, 97.07).

Table 3 suggests that the relative bias, S.D., and MSE are all in good agreement with each other. The relative biases are typically an order of magnitude smaller than the standard deviations, and so the latter dictate the MSEs. One exception is the large *D*-zero simulation for β_3, whose bias is more significant and whose relative MSE is roughly 50% larger than that under EBLUP.

Note that β_3 is the only non-random coefficient, and it enters the model nonlinearly. The relative S.D. and relative MSE increase with the size of the assumed variance components for β_1 and β_2 but remain constant for β_3. Thus, there did not appear to be any "leakage" from the randomness of β_1 and β_2 into the nonrandom β_3. The 95% confidence interval coverages are worst for β_3, with a considerable drop for the large *D*-zero simulation. Also, the coverages tend to decrease from the nominal value for larger variance components.

Our primary conclusion based on Fig. 2 and Table 3 is that both methods provide good results for the fixed-effects parameters assuming small and medium sizes of variance components, but the zero expansion method can potentially be unreliable for large variance components. The EBLUP expansion method displays a considerable improvement in this situation, but the recovery is not complete.

The Medium D-EBLUP results are directly comparable with the LME-RML results of Pinheiro and Bates (1996, Table 6). The bias and RMSE statistics are nearly identical for all three fixed-effects parameters, lending credence to both studies.

5.3. Results for the variance parameters

Fig. 3 contains boxplots for the variance–covariance parameter estimates from the six simulation runs. In these plots, the (D_{11}, D_{21}, D_{22}) values for small D-zero and small D-EBLUP are multiplied by 25 and those for large D-zero and large D-EBLUP are divided by 25 to provide comparability with the medium values.

Fig. 3 suggests that both zero- and EBLUP-expansion methods are highly unstable when the variance components are small in relation to the residual error variance. The resulting estimates of (D_{11}, D_{21}, D_{22}) are highly variable and their distributions are very skewed, especially for D_{22}. When the variance components become larger, both algorithms are more stable and the estimates are less variable. The results for σ^2 remain fairly constant across the simulations and exhibit little bias.

Table 4 contains simulation summary statistics for the variance-covariance parameter estimates. Descriptions of the various statistics are the same as in Table 3. The mean values for D_{11}, D_{21}, and D_{22} change considerably here because the true values are varied for the different simulations. Table 4 suggests that both expansion methods produce very similar results, with the EBLUP expansion improving relative bias only for the large variance components. Somewhat surprisingly, the simulations with small values of the variance components exhibit significantly large bias and variability of the estimates. As with the fixed-effects parameter estimates, the medium D-EBLUP results are very similar to the LME-RML results of Pinheiro and Bates (1996, Table 5).

The 95% confidence intervals for the variance parameters D_{11}, D_{22}, and σ^2 are constructed using a basic Satterthwaite approximation. For example, if the estimate of σ^2 for a particular simulation is denoted $\hat{\sigma}^2$, then the 95% confidence interval for σ^2 is constructed as

$$\left(\frac{\hat{\sigma}^2 v}{\chi^2_{v,0.975}}, \frac{\hat{\sigma}^2 v}{\chi^2_{v,0.025}} \right),$$

where $v = 2[\hat{\sigma}^2/\mathrm{Var}(\hat{\sigma}^2)]^2$ is the approximate degrees of freedom and $\chi^2_{v,\alpha}$ are critical values from the χ^2_v-distribution. The quantity $\mathrm{Var}(\hat{\sigma}^2)$ is the asymptotic sampling variance of $\hat{\sigma}^2$ obtained from the appropriate diagonal element of the inverse of the observed second derivative matrix. The interval for D_{21} is a Wald-type interval constructed by using a Gaussian critical value and asymptotic standard errors obtained in like fashion. For further details on these intervals, see Burdick and Graybill (1992).

As shown in the final column of Table 4, these intervals attain their nominal coverages in a majority of cases, but break down in the small D results for D_{11} and D_{22} and to a lesser extent in the medium D results for D_{22}. The small D cases can be attributed to the large relative bias of the estimates, although the medium D results

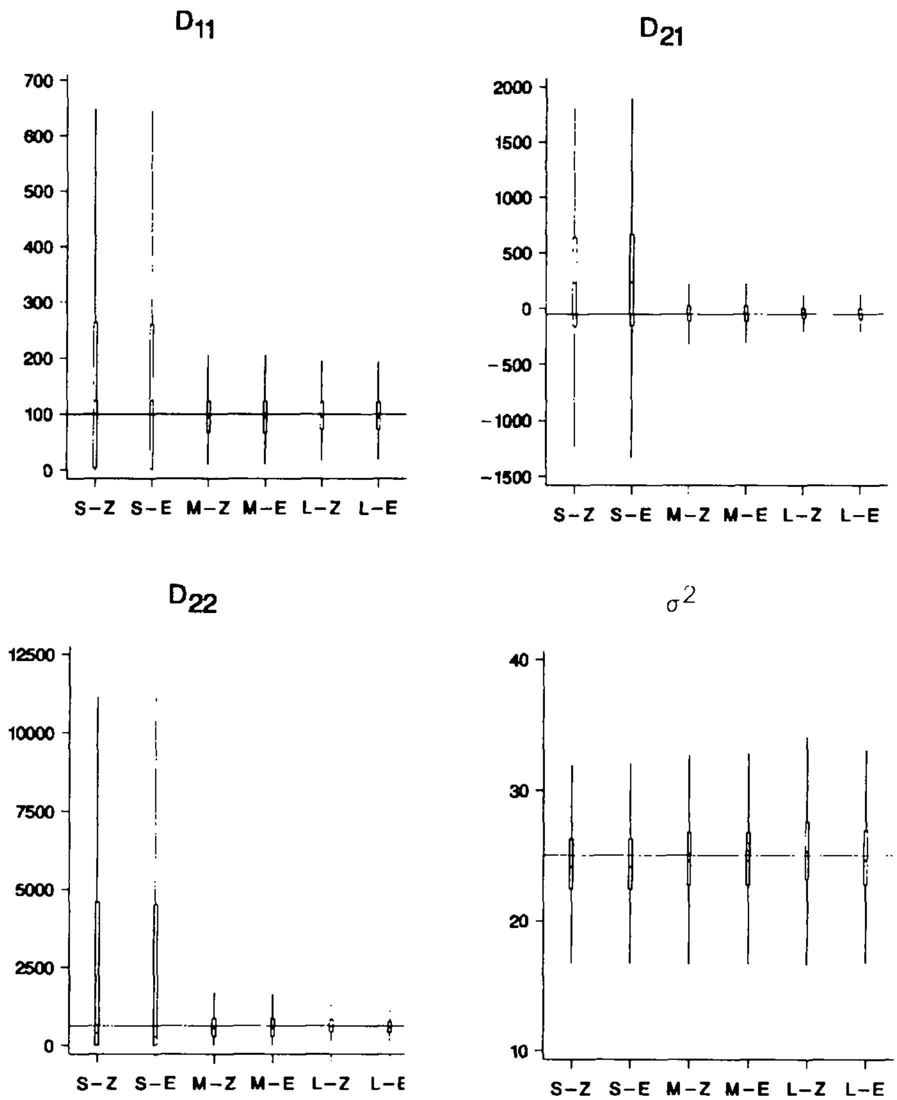

Fig. 3. Box plots of variance–covariance parameter estimates from the simulation settings of the logistic model: S-Z (small D-zero), S-E (small D-EBLUP), M-Z (medium D-zero), M-E (medium D-EBLUP), L-Z (large D-zero), and L-E (large D-EBLUP). The small and large box plots for D_{11}, D_{21}, and D_{22} are scaled to be comparable with the Medium ones, and the horizontal reference lines are located at the true values of the parameters.

for D_{22} undercovered in spite of little relative bias. Finally, the Wald intervals for D_{12} tended to be slightly too conservative for the small and medium D cases.

5.4. Performance and stability

Table 5 displays some performance statistics from the six simulations. Min Iters and Max Iters denote the minimum and maximum number of linear mixed model fits

Table 4
Summary statistics for variance-covariance parameter estimates from the simulations of the logistic model

Simulation	True	Mean	Rel. Bias	Rel. S.D.	Rel. MSD	Rel. RMSE	95% Cov.
D_{11}							
Small D-zero	4	6.9	0.72	1.80	2.36	1.92	52.8*
Small D-EBLUP	4	6.8	0.71	1.78	2.35	1.91	53.3*
Medium D-zero	100	99.7	−0.00	0.46	0.44	0.46	96.3
Medium D-EBLUP	100	99.4	−0.01	0.46	0.44	0.46	96.4
Large D-zero	2500	2489.3	−0.00	0.39	0.38	0.39	94.3
Large D-EBLUP	2500	2461.5	−0.02	0.39	0.38	0.39	94.0
D_{21}							
Small D-zero	−2	14.2	8.20	14.9	17.3	16.9	97.6*
Small D-EBLUP	−2	14.0	8.00	14.7	17.2	16.7	97.8*
Medium D-zero	−50	−43.2	0.14	2.05	2.06	2.05	98.0*
Medium D-EBLUP	−50	−43.8	0.12	2.04	2.05	2.05	98.3*
Large D-zero	−1250	−1023.9	0.18	1.35	1.41	1.36	95.1
Large D-EBLUP	−1250	−1165.0	0.07	1.38	1.38	1.38	95.4
D_{22}							
Small D-zero	25	124.6	3.98	7.12	12.7	8.15	19.8*
Small D-EBLUP	25	123.6	3.95	7.05	12.6	8.08	20.0*
Medium D-zero	625	620.4	−0.01	0.73	0.79	0.73	86.5*
Medium D-EBLUP	625	615.6	−0.02	0.72	0.79	0.72	86.7*
Large D-zero	15625	16550.7	0.06	0.47	0.42	0.48	91.8*
Large D-EBLUP	15625	15230.9	−0.03	0.40	0.39	0.40	95.8
σ^2							
Small D-zero	25	24.4	−0.02	0.13	0.12	0.13	94.4
Small D-EBLUP	25	24.4	−0.02	0.13	0.12	0.13	94.4
Medium D-zero	25	24.8	−0.01	0.13	0.13	0.13	94.9
Medium D-EBLUP	25	24.8	−0.01	0.13	0.13	0.13	94.8
Large D-zero	25	25.5	0.02	0.17	0.13	0.14	94.2
Large D-EBLUP	25	24.9	0.00	0.13	0.13	0.13	94.8

* 95% coverage outside (92.93, 97.07).

Table 5
Performance statistics from the simulations of the logistic model

Simulation	Min Iters	Max Iters	Ave Iters	Ave Sec/It	Non-Conv	% Non-PD D
Small-zero	1	3	1.5	9.8	37	68.7
Small-EBLUP	3	10	5.3	9.2	38	69.1
Medium-zero	1	3	1.4	10.3	3	16.4
Medium-EBLUP	4	10	5.9	9.1	3	16.6
Large-zero	1	6	1.8	8.8	0	0.0
Large-EBLUP	5	13	6.7	9.0	0	0.0

Table 6
Summary statistics for the small-zero simulation conducted while constraining D to be nonnegative definite

Parameter	True	Mean	Rel. Bias	Rel. S.D.	Rel. RMSE
β_1	200	200.0	0.0002	0.01	0.01
β_2	700	700.4	0.0006	0.01	0.01
β_3	350	350.0	−0.0001	0.02	0.02
D_{11}	4	6.4	0.6001	1.71	1.81
D_{21}	−2	16.1	9.0507	12.88	15.74
D_{22}	25	154.0	5.1606	6.68	8.44
σ^2	25	24.4	−0.0246	0.12	0.13

required to attain convergence of the algorithms. Ave Iters is the average number of iterations and Ave Sec/It is the average number of seconds (clock time) per iteration. Non-Conv denotes the number of non-convergences of the algorithm before obtaining 1000 convergences. % non-PD D indicates percentage of the estimates of D which are nonpositive definite.

The iteration statistics in Table 5 indicate that the EBLUP-expansion algorithm requires approximately 4 times more computer work than the zero-expansion. In light of the results in Tables 3 and 4, this extra work appears to be potentially worth the trouble only for large variance components. All of the simulations required approximately 10 s of clock time per iteration. Roughly half of this time is spent in CPU processing and the other half in I/O.

The large amount of variability associated with the small D simulations described previously appears to be associated with some instability in the two algorithms. The 37 and 38 non-convergences of the NLINMIX macro are evidence of this. The non-convergences typically involved estimates of D_{11} or D_{22} near their boundary of zero, and convergence was obtainable for some of them by fiddling with starting values. Other cases came close to the relative convergence criterion of 1×10^{-8}, but hovered around 1×10^{-7} indefinitely.

The small D simulations also exhibit a much larger percentage of estimates of D which are nonpositive definite when compared with the medium and large D studies. Since nearly 70% of the small D simulation estimates of D are nonpositive definite, an additional simulation is carried out to investigate the effect of constraining D to be nonnegative definite. In this extra simulation, D is constrained to be nonnegative definite by parameterizing it in terms of its Cholesky decomposition L, where L is a lower triangular matrix satisfying $D = LL'$. The small D-zero simulation is repeated while enforcing this constraint, and the results from this analysis are summarized in Table 6.

Table 6 reveals basically the same results as Tables 3 and 4, and so the estimates of $(\beta_1, \beta_2, \beta_3)$ are still good and the nonnegative-definite constraint did not improve the bias problems associated with the estimates of (D_{11}, D_{21}, D_{22}). There were 23 nonconvergences and 56.7% of the estimates of D were nonpositive definite, both representing reductions over the unconstrained simulation results of 37 and 68.7%, respectively, from Table 5.

6. One-compartment model simulation

6.1. Model

In this section we simulate observations from the following pharmacokinetic model:

$$y_{ij} = \frac{10k_a[\exp(-k_i t_{ij}) - \exp(-k_a t_{ij})]}{v_i(k_a - k_i)} \times \exp(e_{ij}),$$

where

$$k_i = k_e c_i / v_i,$$

$$c_i = \exp(z_{ci}),$$

$$v_i = \exp(z_{vi}).$$

The observations y_{ij} are simulated concentrations of a drug in the bloodstream of subject i ($i = 1, \ldots, 50$) at time t_{ij} ($j = 1, \ldots, n_i$) and are analyzed on the log scale to provide additive residual errors. The initial times for each subject t_{i1} are randomly selected from a uniform distribution on $(0.1, 0.6)$ and then increase systematically by 1 thereafter. The number of observations for each subject n_i are selected randomly from the discrete uniform distribution on $(1, 7)$; therefore, the design is fairly sparse and unbalanced.

The fixed-effects parameters to be estimated are k_a and k_e. The parameter k_a is the absorption rate constant and k_e is a scaling factor for the elimination rate constant for each subject k_i. The subject-specific constants k_i are assumed to be the ratio of k_e times two lognormal random variables: total body clearance c_i and volume of distribution v_i. The normal random variables z_{ci} and z_{vi} that model c_i and v_i are assumed to be independent with mean zero and variances σ_c^2 and σ_v^2, respectively. The residual errors e_{ij} are assumed to be normally distributed and independent of z_{ci} and z_{vi} with mean zero and variance σ^2. The variance parameters to be estimated are therefore σ_c^2, σ_v^2, and σ^2.

This model is a simplified version of the one-compartment model considered by Roe (1996) and is similar to the first-order compartment model analyzed by Pinheiro and Bates (1996). Note that z_{ci} and z_{vi} enter the model nonlinearly without being added to any fixed-effects parameters. This necessitates the general formulation of the nonlinear mixed model methods presented in Sections 2–4.

For this simulation we use the following settings for the parameters:

$$k_a = 1.0,$$

$$k_e = 0.5,$$

$$\sigma_c^2 = 0.1,$$

$$\sigma_v^2 = 0.1,$$

$$\sigma^2 = 0.001, 0.01, 0.1.$$

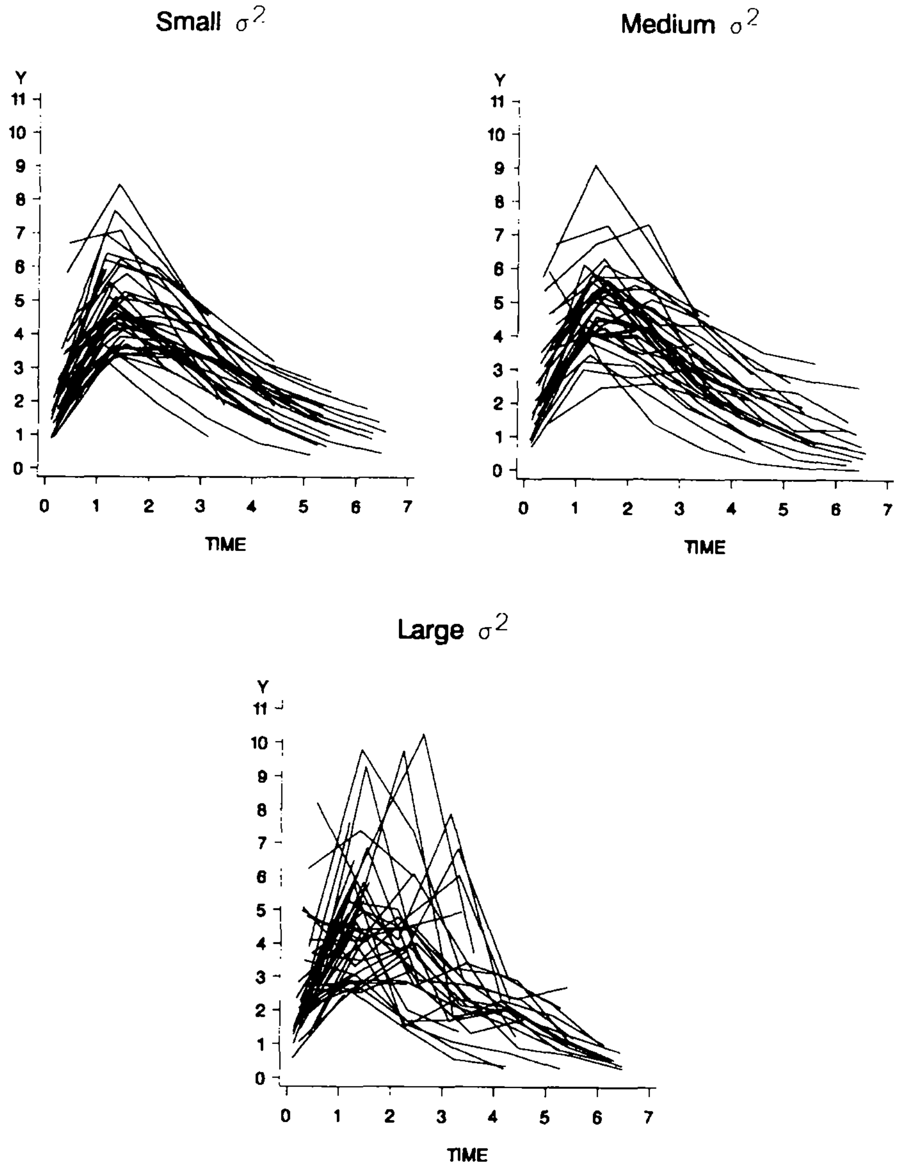

Fig. 4. Example of simulated one-compartment curves for three sizes of residual variance (small, medium, large).

In contrast with the previous simulation for the logistic model, only the residual variance varies in this simulation study. We still use the adjectives small, medium, and large to describe the amount of variability; however, in this study these adjectives refer to the size of σ^2 and not to the variance parameters associated with the random effects. Fig. 4 plots example profiles of the model for the three cases.

As in the logistic curve simulations, 1000 data sets for each of the three cases are generated, and these are each fit using the zero- and EBLUP-expansion methods, resulting in a total of six simulation runs.

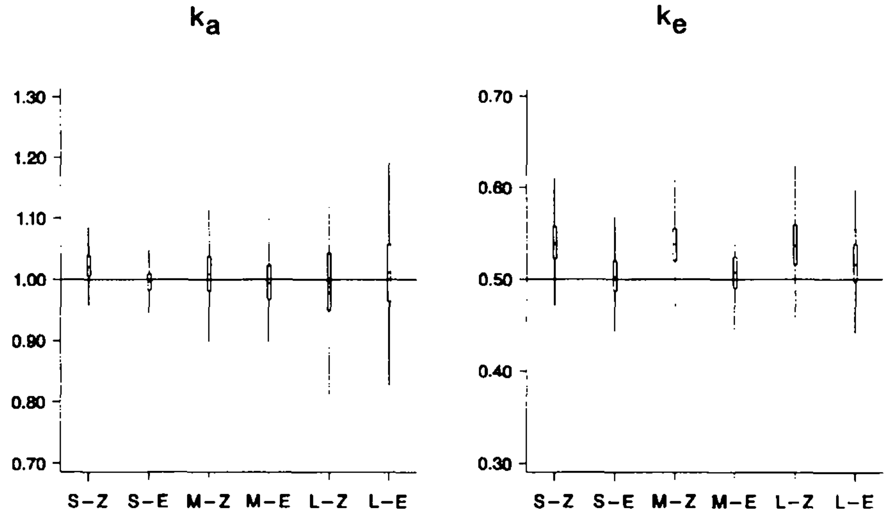

Fig. 5. Box plots of fixed-effects parameter estimates from the simulation settings of the one-compartment model: S-Z (small σ^2-zero), S-E (small σ^2-EBLUP), M-Z (medium σ^2-zero), M-E (medium σ^2-EBLUP), L-Z (large σ^2-zero), and L-E (large σ^2-EBLUP). The horizontal reference lines are located at the true values of the parameters.

Table 7
Summary statistics for fixed-effects parameter estimates from the simulations of the one-compartment model

Simulation	True	Mean	Rel. Bias	Rel. S.D.	Rel. MSD	Rel. RMSE	95% Cov.
k_a							
Small σ^2-zero	1.0	1.022	0.022	0.027	0.035	0.022	79.9*
Small σ^2-EBLUP	1.0	0.995	−0.005	0.021	0.022	0.021	93.9
Medium σ^2-zero	1.0	1.009	0.009	0.045	0.046	0.043	94.0
Medium σ^2-EBLUP	1.0	0.994	−0.006	0.043	0.043	0.042	94.5
Large σ^2-zero	1.0	0.998	−0.002	0.069	0.069	0.068	94.2
Large σ^2-EBLUP	1.0	1.012	0.012	0.067	0.068	0.068	95.6
k_e							
Small σ^2-zero	0.5	0.540	0.079	0.054	0.096	0.053	70.9*
Small σ^2-EBLUP	0.5	0.504	0.008	0.049	0.050	0.050	94.7
Medium σ^2-zero	0.5	0.538	0.076	0.057	0.095	0.056	75.6*
Medium σ^2-EBLUP	0.5	0.508	0.016	0.052	0.054	0.052	95.0
Large σ^2-zero	0.5	0.538	0.076	0.067	0.101	0.065	81.6*
Large σ^2-EBLUP	0.5	0.518	0.036	0.064	0.074	0.063	90.4*

* 95% coverage outside (92.93, 97.07).

6.2. *Results for the fixed-effects parameters*

Fig. 5 and Table 7 display the results for the fixed-effects parameters k_a and k_e. The forms of Fig. 5 and Table 7 are the same those of Fig. 2 and Table 3 discussed in the previous section.

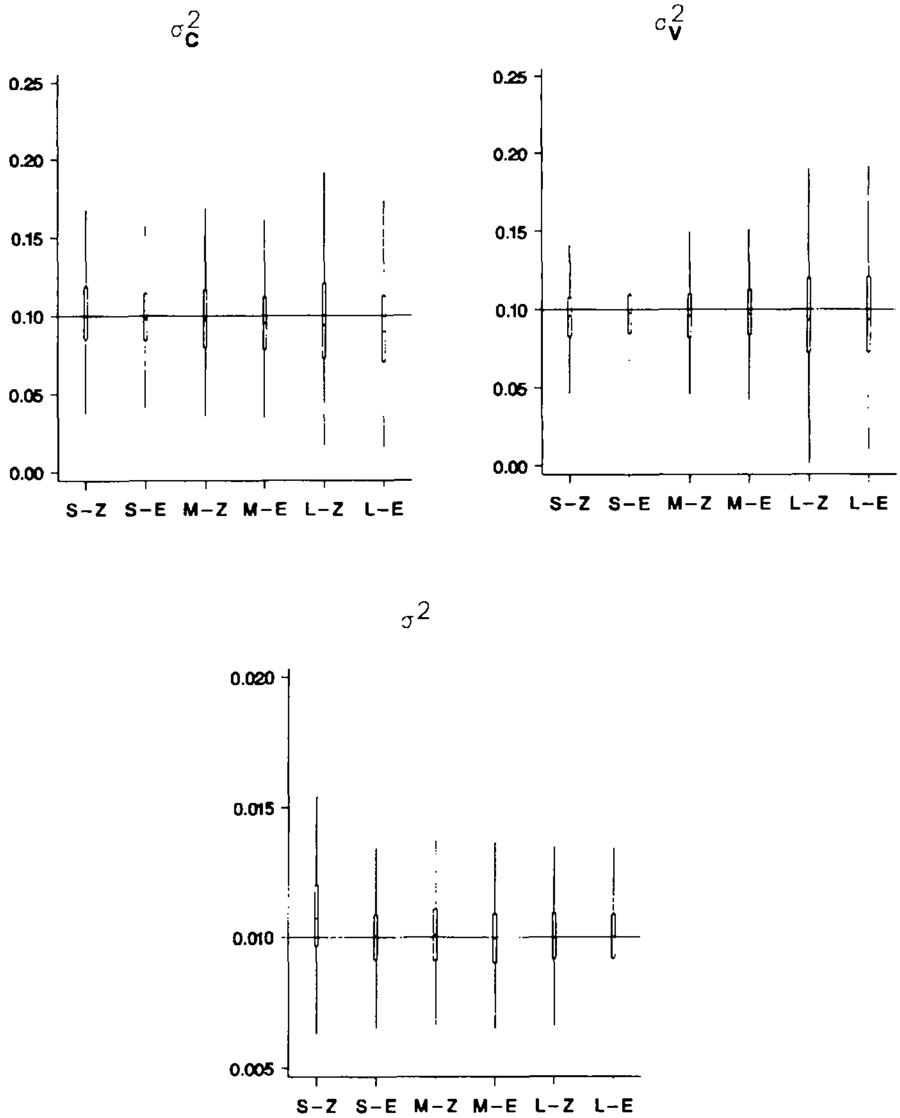

Fig. 6. Box plots of variance parameter estimates from the simulation settings of the one-compartment model: S-Z (small σ^2-zero), S-E (small σ^2-EBLUP), M-Z (medium σ^2-zero), M-E (medium σ^2-EBLUP), L-Z (large σ^2-zero), and L-E (large σ^2-EBLUP). The small and large box plots for σ^2 are scaled to be comparable with the Medium ones, and the horizontal reference lines are located at the true values of the parameters.

The averages of the k_a estimates are very close to target for all simulations except for some upward bias in the small σ^2-zero simulation. However, the EBLUP-expansion method provides noticeable bias improvements over the zero-expansion method in the estimates of k_e for all of the simulations, although bias for it increases with larger residual variances and remains positive.

Table 8
Summary statistics for variance parameter estimates from the simulations of the one-compartment model

Simulation	True	Mean	Rel. Bias	Rel. S.D.	Rel. MSD	Rel. RMSE	95% Cov.
σ_c^2							
Small σ^2-zero	0.100	0.103	0.034	0.264	0.266	0.232	91.9*
Small σ^2-EBLUP	0.100	0.100	0.003	0.229	0.229	0.229	95.0
Medium σ^2-zero	0.100	0.100	0.002	0.272	0.272	0.249	93.1
Medium σ^2-EBLUP	0.100	0.096	−0.036	0.245	0.247	0.246	95.2
Large σ^2-zero	0.100	0.098	−0.017	0.360	0.360	0.323	95.2
Large σ^2-EBLUP	0.100	0.094	−0.065	0.331	0.338	0.322	97.0
σ_v^2							
Small σ^2-zero	0.100	0.096	−0.035	0.188	0.191	0.196	95.5
Small σ^2-EBLUP	0.100	0.098	−0.017	0.190	0.191	0.200	95.8
Medium σ^2-zero	0.100	0.097	−0.025	0.211	0.213	0.215	96.0
Medium σ^2-EBLUP	0.100	0.100	−0.004	0.217	0.217	0.220	96.1
Large σ^2-zero	0.100	0.097	−0.031	0.367	0.369	0.359	96.0
Large σ^2-EBLUP	0.100	0.097	−0.028	0.374	0.374	0.360	95.5
σ^2							
Small σ^2-zero	0.001	0.001	0.124	0.273	0.299	0.155	83.1*
Small σ^2-EBLUP	0.001	0.001	0.005	0.135	0.135	0.139	95.8
Medium σ^2-zero	0.010	0.010	0.013	0.145	0.146	0.138	94.1
Medium σ^2-EBLUP	0.010	0.010	−0.000	0.138	0.138	0.136	95.0
Large σ^2-zero	0.100	0.100	0.003	0.131	0.131	0.132	94.7
Large σ^2-EBLUP	0.100	0.101	0.005	0.130	0.130	0.132	95.0

* 95% coverage outside (92.93, 97.07).

As expected, the estimates for both fixed-effect parameters become more variable with increased residual error. In addition, Table 7 reveals that the average model-based standard deviations exceed the observed standard deviations in the aforementioned cases where bias is observed. However, this variance inflation is not sufficient to provide nominal coverages for the 95% confidence intervals in these cases. Below-normal 95% confidence interval coverages occur in four out of the six zero-expansion results and in the large σ^2-EBLUP results for k_e.

6.3. Results for the variance parameters

Fig. 6 and Table 8 display the results for the variance parameters σ_c^2, σ_v^2, and σ^2. All of the simulations exhibit very little bias and the zero- and EBLUP-expansion methods appear to perform equally well, except for the small σ^2-zero simulation results for σ^2 in which there is some positive bias.

The estimates of σ_v^2 and σ^2 become slightly less precise with larger σ^2 values, and the different measures of dispersion are in close agreement except again for the small σ^2-zero results for σ^2. The Satterthwaite-based confidence intervals maintain their 95% coverages except for the small σ^2-zero results for σ^2 and σ_c^2.

6.4. Performance and stability

The average number of iterations for the zero-expansion simulations were close to 5 for all three sizes of σ^2. For the EBLUP-expansion simulations, the average number of iterations were 9, 13, and 21 for the small, medium, and large sizes, respectively, indicating a decrease in stability of the EBLUP-expansion method for larger residual variances. Computing times were similar to those reported in Table 5. The large σ^2-EBLUP simulation had 27 non-convergences before completing 1000 trials; all other methods had none.

7. Discussion

We would like to make the following remarks about the results of our two simulation studies:

1. Regarding bias in the fixed-effects parameter estimates, the methods perform fairly well in most cases. The EBLUP-expansion method tends to perform better than the zero-expansion method in cases where the ratio D/σ^2 becomes large, that is, in the large D results for the logistic model and the small σ^2 results for the one-compartment model. The zero-expansion method has the most trouble with the β_3 parameter in the logistic model (see Fig. 2) and the k_e parameter in the one-compartment model (see Fig. 5). The EBLUP-expansion method also exhibited some bias with the k_e parameter as σ^2 increased in size.
2. Regarding bias in the variance parameter estimates, the methods do fairly well except for the small D results for the D parameters in the logistic model (see Fig. 3) and the small σ^2 results for σ^2 in the one-compartment model (see Fig. 6).
3. Regarding variability in fixed-effects parameter estimates, the estimates become more variable as more noise is introduced into the models for all cases, agreeing with intuition.
4. Regarding variability in the variance parameter estimates, the relative precision of the parameters that were altered during the simulations (D in the logistic model and σ^2 in the one-compartment model) actually became worse for smaller settings of these parameters (see the D parameter results in Fig. 3 and the σ^2 results in Fig. 6). These small settings also produced the most instability in both the zero- and EBLUP-expansion methods. Finally, the magnitude of the variability for the variance parameter estimates was typically one or more orders larger than the variability of the fixed-effects parameter estimates.

Overall, the EBLUP-expansion estimates are slightly better than the zero-expansion estimates in terms of bias and MSE. These results agree with the previous work of Breslow and Lin (1995), and this provides confirmation of the intuitive notion that the local approximations for each subject offered by the EBLUP expansions do offer some improvement over the mean-value based approximations of the Zero expansions. However, the magnitude of the improvement is difficult to predict from the analytical results. For our examples the improvement is slight, although this certainly may not be true for other nonlinear models.

One noticeable contrast in our two simulation studies is that for the logistic model the estimates of the fixed-effects parameters remain nearly unbiased even when the variance components are poorly estimated, but for the one-compartment model biases are prevalent even for accurate estimates of the variance parameters. One source of explanation for this discrepancy is the principle of orthogonality between the fixed-effects and variance parameters; see Barndorff-Nielsen and Cox (1989) and Wolfinger (1993). No additional discussion is presented here except to remark that in cases where orthogonality is severely violated, linearization methods tend to break down (for example, generalized linear mixed models for binary data as discussed in Breslow and Clayton, 1993). Further work is needed to determine the extent to which nonorthogonality affects the general applicability of the approximation methods.

Finally, the results of Shun and McCullagh (1995) and Vonesh (1996) reveal that the bias of Laplace approximations similar to (23) may not vanish as the sample size increases. If the number of observations with each subject is small and the random effects enter into the model in a nonlinear fashion, both estimation procedures may not perform well, even when the number of subjects is large. We therefore recommend that analysts conduct simulation studies to evaluate linearization methods for models and data sets dissimilar to the ones we consider in Sections 5 and 6.

References

Barndorff-Nielsen, O.E., Cox, D.R., 1989. Asymptotic Techniques for Use in Statistics. Chapman & Hall, London.

Bates, D.M., Watts, D.G., 1980. Relative curvature measures of nonlinearity (with discussion). J. Roy. Statist. Soc. Ser. B 42, 1–25.

Beal, S.L., Sheiner, L.B. (Ed.), 1992. NONMEM User's Guide. University of California, San Francisco, NONMEM Project Group.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc. 88, 9–25.

Breslow, N.E., Lin, X., 1995. Bias correction in generalized linear mixed models with a single component of dispersion. Biometrika 82, 81–91.

Burdick, R.K., Graybill, F.A., 1992. Confidence Intervals on Variance Components. Marcel Dekker, New York.

Davidian, M., Gallant, A.R., 1993. The nonlinear mixed effects model with a smooth random effects density. Biometrika 80, 475–488.

Engel, B., Keen, A., 1994. A simple approach for the analysis of generalized linear mixed models. Statist. Neerlandica 48, 1–22.

Goldstein, H., 1991. Nonlinear multilevel models, with an application to discrete response data. Biometrika 78, 45–51.

Green, P.J., 1987. Penalized likelihood for general semi-parametric models. International Statistical Review 55, 245–259.

Gumpertz, M., Pantula, S.G., 1992. Nonlinear regression with variance components. J. Amer. Statist. Assoc. 87, 201–209.

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Amer. Statist. Assoc. 72, 320–340.

Harville, D.A., 1990. BLUP (best linear unbiased prediction), and beyond. In Advances in Statistical Methods for Genetic Improvement of Livestock. Springer, New York, pp. 239–276.

Henderson, C.R., 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Canada.

Hirst, K., Zerbe, G.O., Boyle, D.W., Wilkening, R.B., 1991. On nonlinear random effects models for repeated measurements. Comm. Statist. Simulation 20. 463–478.

Kackar, R.N., Harville, D.A., 1984. Approximations for standard errors of estimators of mixed and random effects in mixed linear models. J. Amer. Satist. Assoc. 79, 853–862.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.

Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

Lindstrom, M.J., Bates, D.M., 1990. Nonlinear mixed effects models for repeated measures data Biometrics 46, 673–687.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., 1996. SAS System for Mixed Models. SAS Institute Inc., Cary, NC.

Longford, N., 1988. A quasi-likelihood adaptation for variance component analysis. Proc. on Computational Statistics. American Statistical Association. Alexandria, VA, pp. 137–142.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed. Chapman & Hall, London.

McGilchrist, C.A., 1994. Estimation in generalized linear mixed models. J. Roy. Statist. Soc. Ser. B 56, 61–69.

Pinheiro, J.C., Bates, D.M., 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. J. Comput. Graphical Statist. 4. 12–35.

Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects (with discussion). Statist. Sci. 6, 15–51.

Roe, D.J., 1996. Comparison of population pharmacokinetic modeling methods using simulated data: results from the population modeling workgroup. Statist. Med., to appear.

SAS Institute Inc., 1996. SAS/STAT Software: Changes and Enhancements through Release 6.11. SAS Institute Inc., Cary, NC.

Schall, R., 1991. Estimation in generalized linear models with random effects. Biometrika 78, 719–727.

Sheiner, L.B., Beal, S.L., 1980. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis–Menten model: routine clinical pharmacokinetic data. J. Pharmacokinetics Biopharmaceutics 8, 553–571.

Shun, Z., McCullagh, P., 1995. Laplace approximation of high dimensional integrals. J. Roy. Statist. Soc. Ser. B 57, 749–760.

Solomon, P.J., Cox, D.R., 1992. Nonlinear components of variance models. Biometrika 79, 1–11.

Tempelman, R.J., Gianola, D., 1995. A mixed effects model for overdispersed count data in animal breeding. Biometrics 52, 265–279.

Tierney, L., Kadane, J.B., 1986. Accurate approximations for posterior moments and marginal densities. J. Amer. Statist. Assoc. 81, 82–86.

Vonesh, E.F., 1992. Nonlinear models for the analysis of longitudinal data, Statist. Med. 11, 1929–1954.

Vonesh, E.F., 1996. A note on Laplace's approximation in nonlinear mixed effects models. Biometrika 83, 447–452.

Vonesh, E.F., Carter, R.L., 1992. Mixed-effects nonlinear regression for unbalanced repeated measures. Biometrics 48, 1–17.

Wakefield, J., 1996. The Bayesian analysis of population pharmacokinetic models. J. Amer. Statist. Assoc. 91, 62–75.

Wolfinger, R.D., 1993. Laplace's approximation for nonlinear mixed models. Biometrika 80, 791–795.

Wolfinger, R.D., O'Connell, M., 1993. Generalized linear mixed models: a pseudo-likelihood approach. J. Statist. Comput. Simulation 48, 233–243.

Zeger, S.L., Karim, M.R., 1991. Generalized linear models with random effects: a Gibbs sampling approach. J. Amer. Statist. Assoc. 86, 79–86.

Zeger, S.L., Liang, K.Y., 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42, 121–130.

Zeger, S.L., Liang, K.Y., Albert, P.S., 1988. Models for longitudinal data: a generalized estimating equation approach. Biometrics 44, 1049–1060.