Asymptotic Properties of Nonlinear Mixed-Effects Models in Large Samples

Eugene Demidenko

Dartmouth Medical School, HB 7927
DHMC, Lebanon, NH 03756
e-mail: eugene.demidenko@dartmouth.edu

November 19, 1998

A short version of this paper entitled
"Asymptotic Properties of Nonlinear Mixed-Effects Models"
has been published in

Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions.
Springer Lectures in Statistics. New York: Springer-Verlag, 1997, pp.49-62.

ABSTRACT. The asymptotic properties of four estimators for nonlinear mixed-effects models are investigated: maximum likelihood estimator (MLE), an estimator based on the first-order approximation to the expectation function (Vonesh and Carter (1992)), the two-stage, and the Lindstrom-Bates (1990) estimators. Two asymptotic situations are considered: (i) when $N \to \infty$ and n_i is finite, (ii) $N \to \infty$ and $\min n_i \to \infty$; where N is the number of individuals and n_i is the number of observations on the ith individual. For a simple one-parameter balanced exponential model only the MLE is consistent when the numbers of observations per individuals, $\{n_i\}$ are finite. The estimator based on the first-order approximation is always inconsistent, i.e., has a systematic bias. The asymptotic bias for the other three estimators is evaluated when $N \to \infty$ and $n_i = \text{const}$. When $N \to \infty$ and $\min n_i \to \infty$; the MLE, Lindstrom-Bates and two-stage estimators are equivalent. The bias for the two-stage estimator based on a second order approximation is evaluated in general nonlinear mixed-effects model, and a bias-corrected version of this estimator is proposed.

Key words and phrases: Random effects, Maximum likelihood, Repeated measures, Longitudinal data.

1. Introduction

Nonlinear mixed-effects (NLME) model is a natural generalization of the linear mixed-effects model introduced by Laird and Ware (1982). The NLME model has received substantial attention in the literature; recently published monographs by Davidian and Giltinan (1995), and Vonesh and Chinchilli (1997) provide a comprehensive source of up-to-date knowledge on the subject. However, until now its asymptotic properties have not be investigated in detail. It is well known that maximum likelihood estimation for nonlinear mixed-effects models leads to a cumbersome integration problem because random parameters appear inside the nonlinear expectation function. To avoid this problem several other estimators have been proposed. For instance, Pocock et al. (1981), Berkey and Laird (1986) suggested using an intuitively appealing two-stage estimator. Another approach based on the first-order approximation to extract the random effects from the expectation regression function was originally proposed by Sheiner and Beal (1980) and developed by Vonesh and Carter (1992). Another method to use the first-order approximation was suggested by Lindstrom and Bates (1990), however they approximate nonlinear functions around estimates of random effects. Wolfinger (1993) showed that the Lindstrom-Bates estimator can be obtained using Laplace's approximation to the likelihood function. Recently, Pinheiro and Bates (1995) compared several estimators for nonlinear mixed-effects models via statistical simulations and suggested some numerical hints. Two recently published monographs on NLME models by Davidian and Giltinan (1995), and Vonesh and Chinchilli (1997) provide a comprehensive source of up-to-date knowledge in this area.

Little is known about the statistical properties of estimators for nonlinear mixed-effects models even in large samples and this question is open (e.g., Davidian and Giltinan, 1995). Unlike linear models, it is impossible to state any general statistical properties for nonlinear models in finite samples even when variance parameters are known. The only way to make any general statements on the NLME model estimation is to consider asymptotic properties of estimators, i.e., when the number of observations goes to infinity. In mixed-effects models the following two asymptotic situations should be distinguished:

1. The number of individuals, N tends to infinity and the number of observations per individual, n_i remains finite, i.e. uniformly bounded.

2. The number of individuals N tends to infinity along with $\min n_i$.

For linear models the first condition is sufficient to guarantee the consistency, asymptotic normality and efficiency (Vonesh and Carter 1987, Stukel and Demidenko 1996a). For instance, Vonesh and Carter (1992) proved that their four-step estimator is normally distributed and asymptotically efficient when N tends to infinity along with $\min_i n_i$ if the original model is replaced by its first-order approximation, 'pseudo-model'. However, it is important to realize that the approximate pseudo-model does not coincide with the original one, and their results cannot be extended to the originally formulated nonlinear model with random effects inside the nonlinear regression function. Recently, Vonesh (1996) provided some arguments for consistency of Lindstrom-Bates types of estimators based on the Laplace's approximation of the log-likelihood function when N tends to infinity along with $\min n_i$.

The goal of the present paper is to establish asymptotic properties of the above estimators in the two asymptotic situations specified above.

The structure of the paper is as follows. In section 2 the general nonlinear mixed-effects model and the four estimators are introduced: the maximum likelihood estimator, the Vonesh-Carter, the two-stage and the Lindstrom-Bates estimators. In section 3 a simple one-parameter exponential nonlinear mixed-effects model is investigated. An attractive feature of this model is that the first-order approximation and the two-stage estimators admit a closed form solution. The exact asymptotic bias when $N \to \infty$ and $n_i = n = \text{const}$ is calculated for all estimators. In section 4 we demonstrate the equivalence of the ML, two-stage and Lindstrom-Bates estimators in the first asymptotic situation. A bias-corrected version of the two-stage estimator is proposed in section 5.

2. The NLME model and four estimators

In this section the general nonlinear mixed-effects model is formulated and four estimators are introduced: the maximum likelihood, Vonesh-Carter, the two-stage and Lindstrom-Bates estimators.

2.1. The NLME model

The general nonlinear mixed-effects model is written in two stages (Sheiner and Beal 1985, Lindstrom and Bates 1990, Davidian and Giltiman 1995, Vonesh and Chinchilli 1997). The first stage consists of N within-individual nonlinear regression models with random parameters

$$y_i = f_i(a_i) + \varepsilon_i; \quad i = 1, ..., N \tag{2.1}$$

where: y_i is $n_i \times 1$ vector dependent variable, f_i is a nonlinear $n_i \times 1$ vector function, $f_i(a_i) = (f(a_i; x_{i1}), f(a_i; x_{i2}), ..., f(a_i; x_{i;n_i}))'$ and x_{ij} is a vector of covariates.; a_i is an unobservable $m \times 1$ vector of random parameters with unknown $m \times m$ covariance matrix $\sigma^2\Omega$; ε_i is the $n_i \times 1$ error vector with $E(\varepsilon_i) = 0$ and $\mathrm{cov}(\varepsilon_i) = \sigma^2 I$; N is the number of individuals, n_i is the number of observations on the ith individual. The second stage is a linear model

$$a_i = Z_i\beta + b_i \tag{2.2}$$

where Z_i is the $m \times k$ design matrix, and β is the $k \times 1$ vector of population parameters of interest; b_i is the random effect, $E(b_i) = 0$; $\mathrm{cov}(b_i) = \sigma^2\Omega$. The random effect b_i and common error term ε_i are assumed independent to each other and between individuals. Also, it is assumed matrix $\sum Z_i'Z_i$ has full rank, $\sigma^2 > 0$ and matrix Ω is positive definite. Vector $\theta = (\sigma^2, \mathrm{vech}'(\Omega))'$ has dimension $1 + m(m+1)/2$ and is called the variance parameter. Also, the random terms ε_i and b_i are assumed to have normal distributions. Thus, the NLME model can be rewritten as

$$y_i \mid a_i \sim N\left(f_i(a_i), \sigma^2 I\right); \tag{2.3}$$

$$a_i \sim N\left(Z_i\beta, \sigma^2\Omega\right) \tag{2.4}$$

with the log-likelihood function up to a constant term

$$l(\beta; \theta) = -\frac{N}{2}\log|\Omega| - \frac{\sum(n_i + m)}{2}\log(\sigma^2) + \sum_{i=1}^{N}\log\int_{R^m} g_i(a; \beta, \theta)da \tag{2.5}$$

where

$$g_i(a; \beta, \theta) = \exp\left(-\frac{1}{2\sigma^2}\|y_i - f_i(a)\|^2\right)\exp\left(-\frac{1}{2\sigma^2}(a - Z_i\beta)'\Omega^{-1}(a - Z_i\beta)\right).$$

Four estimators of β are considered in the following subsections. They all coincide for the linear mixed-effects model with known variance parameters. However, they are all different and have quite different statistical properties applied to the general NLME model.

2.2. Maximum likelihood estimator

Unlike linear mixed-effects model there is no closed form solution to the MLE of NLME model even when variance parameters σ^2 and Ω are known. Moreover, the presence of an integral in (2.5) makes the maximum likelihood procedure cumbersome especially for multidimensional random effects.

As follows from maximum likelihood theory, the MLE is consistent and asymptotically normally distributed and efficient. Thus, the asymptotic variance based on the information matrix is the milestone for efficiency comparisons to other estimators. Recently such comparisons has been done for logistic regression model with random intercept term by Neuhaus and Lesperance (1996). Unfortunately, the calculation of the information matrix for NLME models again leads to an integration problem. However, for NLME model it is possible to come up with a lower bound for the variance of the beta-estimates. Thus, our current aim is to find such a bound for the asymptotic variance of the MLE. We use the following result which provides an upper bound for the information matrix in a model with two random vector variables Y and A; where A is unobservable. It is well known that the information based on Y is less than the information based on A if Y is a function of A (e.g., Schervish 1995). We generalize this result in the case when Y depends on A stochastically.

Lemma 2.1. Let Y and A be two random vector variables, A is unobservable with marginal density $f_2(a;\tau)$ where τ is parameter vector. Variable Y is observable and the conditional density of $Y|A$ is $f_1(y;a)$: Therefore the marginal density for Y is $f(y;\tau) = \int f_1(y;a)f_2(a;\tau)da$: Then,

$$J_Y \leq J_A; \tag{2.6}$$

i.e., matrix $J_A - J_Y$ is positive semidefinite, where J_Y is the Fisher information about τ based on Y and J_A is the information based on A:

Applying this Lemma to the model (2.3), (2.4) we come to the following absolute lower bound for the asymptotic covariance of the MLE:

$$\mathrm{cov}(\widehat{\beta}_{ML}) \geq \sigma^2 \left(\sum_{i=1}^{N} Z_i' \Omega^{-1} Z_i \right)^{-1} : \tag{2.7}$$

We notice that this inequality remains true whether or not the variance parameters are known. Also, it is interesting to note that the lower bound (2.7) depends neither on the first stage model nor on the number of observations per individual, n_i: As it is follows from section 5 this inequality turns into equality when $N \to \infty$ and $\min n_i \to \infty$:

Formula (2.7) has an interesting interpretation: it corresponds to a hypothetical situation when the a_i are observable. Indeed, in that case the best estimator, as follows from (2.4), would be a generalized least squares estimator with the covariance matrix coinciding with the right-hand side of (2.7).

2.3. Vonesh-Carter estimator

To circumvent the integration problem, Vonesh and Carter (1992) suggested to replace the NLME model by another nonlinear model with linear random effects derived as the first-order approximation to (2.1) around $E(a_i) = Z_i\beta$. Thus, instead of (2.1, 2.2) a pseudo-model is considered:

$$y_i = f_i(Z_i\beta) + R_i(Z_i\beta)b_i + \varepsilon_i; \tag{2.8}$$

where $R_i = \partial f_i / \partial a_i$ and $b_i \sim N(0, \sigma^2 \Omega)$: In particular, a four-step estimation procedure based on (2.8) was suggested by Vonesh and Carter when the variance parameters are unknown. They proved that for the pseudo-model (2.8), the four-step (VC) estimator is asymptotically equivalent to the MLE when both $N \to \infty$ and $\min n_i \to \infty$: However, it is important to note that the model (2.8) does not coincide with the original model (2.1, 2.2) and, consequently, the properties of this estimator do not necessarily remain true for the original model. Moreover, as it is shown in section 3, the VC-estimator is not consistent for model (2.1) in any of two asymptotic situations.

2.4. Two-stage estimator

Several authors (Pocock et al. 1981, Berkey and Laird 1986) proposed a two-stage (TS) procedure to estimate parameters β: (i) estimate a_i individually by nonlinear least squares, (ii) substitute them into the second stage model and apply generalized least squares. At the first step, N individual nonlinear regression problems are solved using standard techniques:

$$\| y_i - f_i(\alpha_i) \|^2 \Rightarrow \min_{\alpha_i}; \qquad i = 1, \ldots, N \tag{2.9}$$

to obtain $\widehat{a}_i^{LS}$: The conditional covariance matrix for $\widehat{a}_i^{LS}$, based on the standard theory of nonlinear regression, is $V_i \equiv \operatorname{cov}(\widehat{a}_i^{LS} \mid b_i) = \widehat{\sigma}^2 (R_i' R_i)^{-1}$; where R_i is the $n_i \times m$ derivative matrix of f_i at $\widehat{a}_i^{LS}$ and $\widehat{\sigma}^2$ is the pooled variance estimate. Noting that

$$\operatorname{cov}(\widehat{a}_i^{LS}) = \operatorname{cov}(\widehat{a}_i^{LS} \mid b_i) + \operatorname{cov}(b_i) = V_i + \sigma^2 \Omega; \tag{2.10}$$

the generalized least squares is applied to (2.2) to obtain the two-stage (TS) estimator,

$$\widehat{\beta}_{TS} = \left[\sum Z_i' (V_i + \widehat{D})^{-1} Z_i \right]^{-1} \left[\sum Z_i' (V_i + \widehat{D})^{-1} \widehat{a}_i^{LS} \right]: \tag{2.11}$$

If the estimation of (2.9) fails for some i due to nonconvergence we do not use it in the sums in (2.11). However, the number of nonfailed individual regressions must be large enough to allow inverse the matrix in (2.11). An estimate for Ω can be obtained as the MLE from the second stage model substituting $\widehat{a}_i^{LS}$ instead of a_i in (2.4), the global two-stage method (Davidian and Giltiman 1995, Stukel and Demidenko 1997), or as $\widehat{D} = \widehat{D}_1 - \widehat{D}_2$ where

$$\widehat{D}_1 = \frac{1}{N} \sum d_i d_i'; \quad \widehat{D}_2 = \frac{1}{N} \sum V_i \tag{2.12}$$

and $d_i = \widehat{a}_i^{LS} - Z_i (\sum Z_j' Z_j)^{-1} (\sum Z_j' \widehat{a}_j^{LS})$ is the OLS-residual vector for (2.2). As it was recently shown by Stukel and Demidenko (1997) the TS-estimator for linear mixed-effects model is consistent, asymptotically normally distributed and efficient even in a more general setting when $N \to \infty$ and $\{n_i\}$ are fixed.

2.5. Lindstrom-Bates estimator

Lindstrom and Bates (1990) suggested to approximate the nonlinear function f_i not around $E(a_i) = Z_i \beta$; as Vonesh and Carter did, but around $E(a_i) + \widehat{b}_i$ where $\widehat{b}_i$ is an 'estimate' of b_i: The suggested procedure consists of two steps:

1. Penalized nonlinear least squares. Given Ω the following minimization problem is solved:

$$\sum_{i=1}^{N} \left[\| y_i - f_i(Z_i\beta + \tau_i) \|^2 + \tau_i' \Omega^{-1} \tau_i \right] \Rightarrow \min_{\beta, \tau_1, \dots, \tau_N} .$$

2. Linear mixed-effects. Given 'estimates' of the random effects $\hat{b}_i$ and $\hat{\beta}$, derived at the previous step, apply the linear mixed-effects maximum likelihood estimation procedure based on the model

$$w_i \sim N\left(R_i Z_i \beta, \sigma^2 (I + R_i \Omega R_i') \right) \tag{2.13}$$

with pseudo-observations w:

$$w_i = y_i - f_i(Z_i\hat{\beta} + \hat{b}_i) + R_i\hat{b}_i + R_i Z_i \hat{\beta}, \quad R_i = \left. \frac{\partial f_i}{\partial a_i} \right|_{a_i = Z_i\hat{\beta} + \hat{b}_i}$$

to find next estimates $\hat{\beta}, \hat{\sigma}^2$ and $\hat{\Omega}$.

The algorithm alternates between these two steps until convergence: Wolfinger (1993) pointed out that the two steps agree at convergence. He also showed that this procedure can be derived by applying Laplace's approximation to the log-likelihood function (2.5); see also Vonesh (1996, 1997).

3. The one-parameter exponential model

In this section a simple example of a nonlinear mixed-effects model is considered, a one-parameter exponential model with random parameter. The simplicity of this model allows us to find a closed form solution and calculate the bias directly or find a good approximation for all four estimators considered in the previous section. On the other hand, this model reflects all features of more complicated nonlinear mixed-effects models like bias or possible nonexistence of estimates experienced in practice (Demidenko 1995). Based on this model, we show that all estimators, except the MLE, are inconsistent when $\{n_i\}$ are finite and $N \to \infty$. In the next section this model will be applied to illustrate the asymptotic mean square error (MSE) calculation and efficiency comparison.

The following one-parameter balanced exponential model is considered

$$y_{ij} = e^{a_i} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2), \qquad i = 1, \dots, N;\ j = 1, \dots, n \tag{3.1}$$

with the second stage model

$$a_i = \beta + b_i \quad \delta_i \sim N(0, \sigma^2 \omega^2), \tag{3.2}$$

where the ε_{ij} and b_i are independent, only y_{ij} is observable. It implies that $\bar{y}_i = \sum y_{ij}/n$, $i = 1, \dots, N$ are independent and identically distributed (iid). To simplify, only β, the parameter of interest, is assumed to be unknown in this section, i.e., parameters $\sigma^2 > 0$ and $\omega^2 > 0$ are known. In this section we determine the asymptotic bias, as a function of n, for three estimators when $N \to \infty$ and $n_i = n$ are fixed. It should be noted that for finite n some estimators may not exist with a nonzero probability even for large N. Therefore, we have to take this into account when comparing estimators.

3.1. The maximum likelihood estimator

The likelihood function for the model defined by (3.1) and (3.2) for the ith individual is

$$l_i(\beta; y_{i1}, \ldots, y_{in}) = C(2\pi\sigma^2)^{-n/2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_{ij}-e^a)^2} e^{-\frac{1}{2\omega^2\sigma^2}(a-\beta)^2} da$$

where $C = (2\pi\sigma^2\omega^2)^{-1/2}$; with the information

$$\begin{aligned} I_\beta &= \int_{R^n} \frac{(dl/d\beta)^2}{l} dy_1 \cdots dy_n = \frac{C}{\sigma^4\omega^4}(2\pi\sigma^2)^{-n/2} \int_{R^n} \frac{g_1^2(\sum_1^n y_j)}{g_2(\sum_1^n y_j)} e^{-\frac{1}{2\sigma^2}\sum_1^n y_j^2} dy_1 \cdots dy_n \\ &= \frac{C}{\sigma^4\omega^4} E_{\overline{y}}\left(\frac{g_1^2(\overline{y})}{g_2(\overline{y})}\right) = \frac{\sqrt{n}}{(2\pi)^{3/2}\sigma^7\omega^6} \int_{-\infty}^{\infty} \frac{g_1^2(s)}{g_2(s)} \exp\left(-\frac{ns^2}{2\sigma^2}\right) ds; \end{aligned} \tag{3.3}$$

where

$$\begin{aligned} g_1(s) &= \int_{-\infty}^{+\infty}(a-\beta)p(a;s)da; \quad g_2(s) = \int_{-\infty}^{+\infty} p(a;s)da \\ p(a;s) &= \exp\left(n\frac{2e^a s - e^{2a}}{2\sigma^2} - \frac{(a-\beta)^2}{2\omega^2\sigma^2}\right). \end{aligned}$$

The asymptotic variance $\mathrm{var}_{as}(\sqrt{N}\widehat{\beta}_{ML}) = I_\beta^{-1} \geq \sigma^2\omega^2$; as follows from (2.7). As we see, there is no closed form solution for the MLE even for this simple model. Recall, that $\widehat{\beta}_{ML}$ is consistent and asymptotically normally distributed when $N \to \infty$ and n is fixed. As it will be shown later, the asymptotic variance of the MLE attains its absolute lower bound $\sigma^2\omega^2$ when $n \to \infty$. We shall use formula (3.3) in section 5 when comparing the MSE of different estimates.

3.2. Vonesh-Carter estimator

Following the idea of Vonesh and Carter we substitute (3.1), (3.2) by, as they called, "pseudo-model"

$$y_{ij} = e^\beta + \eta_{ij} \tag{3.4}$$

where $\eta_{ij} = e^\beta b_i + \varepsilon_{ij}$. Letting $\eta_i = (\eta_{i1}, \eta_{i2}, \ldots, \eta_{in})'$ and $1 = (1, 1, \ldots, 1)'$; the covariance matrix for the random vector term η_i can be written as $\mathrm{cov}(\eta_i) = \sigma^2 V(\beta)$ where $V(\beta) = I + e^{2\beta}\omega^2 11'$. Let an initial estimate of β be given, β_*. Then the VC-estimator is the solution to the following nonlinear regression problem with the weight matrix $V_* = V(\beta_*)$:

$$\sum_{i=1}^{N}(y_i - e^\beta 1)' V_*^{-1}(y_i - e^\beta 1) \Rightarrow \min_\beta.$$

Clearly, the solution is

$$\widehat{\beta}_{VC} = \ln\left(\frac{\sum_i 1'V_*^{-1}y_i}{N\,1'V_*^{-1}1}\right). \tag{3.5}$$

The special structure of V_* allows us to write down its inverse

$$V_*^{-1} = I - \frac{e^{2\beta_*}\omega^2}{1 + e^{2\beta_*}\omega^2 n} 11'.$$

Then, the VC-estimator is simplified as follows:

$$\hat{\beta}_{VC} = \ln\left(\frac{1}{Nn}\sum_{i,j} y_{ij}\right). \tag{3.6}$$

It is interesting to notice that for this model the VC-estimator does not depend on the initial estimate β_*. Now we calculate the asymptotic limit of (3.6). We have

$$\begin{aligned}
\frac{1}{Nn}\sum_{i,j} y_{ij} &= \frac{1}{N}\sum_i\left(\frac{1}{n}\sum_j\left(e^{\beta+b_i}+\varepsilon_{ij}\right)\right) \\
&= \frac{1}{N}\sum_i\left(e^{\beta+b_i}+\frac{1}{n}\sum_j \varepsilon_{ij}\right) = \frac{1}{N}\sum_i e^{\beta+b_i}+\frac{1}{Nn}\sum_{i,j}\varepsilon_{ij}.
\end{aligned}$$

But $e^{\beta+b_i}$ has a lognormal distribution that implies $Ee^{\beta+b_i} = e^{\beta+\sigma^2\omega^2/2}$ (e.g., Rao 1973). Thus, $E(y_{ij}) = E\left(e^{\beta+b_i}+\varepsilon_{ij}\right) = e^{\beta+\sigma^2\omega^2/2}$, and by the law of large numbers,

$$\lim_{N\to\infty}\frac{1}{N}\sum_i e^{\beta+b_i} = e^{\beta+\sigma^2\omega^2/2}, \quad \lim_{N\to\infty}\lim_{n\to\infty}\frac{1}{Nn}\sum_{i,j}\varepsilon_{ij} = 0.$$

Combining these two limits, we obtain

$$\lim_{N\to\infty}\hat{\beta}_{VC} = \ln e^{\beta+\sigma^2\omega^2/2} = \beta+\frac{1}{2}\sigma^2\omega^2. \tag{3.7}$$

Therefore, the VC-estimator is inconsistent and has a systematic positive bias even if $\min n_i \to \infty$ when $N\to\infty$.

The probability that the VC-estimator exists goes to 1 when $N\to\infty$ regardless of n because $\sum y_{ij}/Nn$ converges to $e^{\beta+\sigma^2\omega^2/2}>0$ with probability 1.

3.3. Two-stage estimator

The individual least squares estimation of (3.1) leads to $\hat{a}_i = \ln\bar{y}_i$. Given 'estimate' of a_i, one can apply generalized least squares and come to the TS-estimator,

$$\hat{\beta}_{TS} = \frac{1}{N}\sum_{i=1}^{N}\ln\bar{y}_i. \tag{3.8}$$

Again, by the law of large numbers

$$\lim_{N\to\infty}\hat{\beta}_{TS} = E\ln(\bar{y}_1) = E\ln\left(e^{\beta+b_1}+\bar{\varepsilon}_1\right), \tag{3.9}$$

with probability 1, where

$$\bar{\varepsilon}_1 = \frac{1}{n}\sum_{j=1}^{n}\varepsilon_{1j} \sim N\left(0,\frac{\sigma^2}{n}\right),$$

$b_1 \sim N(0,\sigma^2\omega^2)$ and $\bar{\varepsilon}_1$ are independent.

The TS-estimator has a systematic negative bias for fixed n. It follows from the inequality $\ln(a + x) < \ln(a) + x/a$ for $x \neq 0, a > 0, a + x > 0$ and the fact that b_1 and $\bar{\varepsilon}_1$ are independent:

$$
\begin{aligned}
\lim_{N\to\infty} \hat{\beta}_{TS} &= E \ln\left(e^{\beta+b_1} + \bar{\varepsilon}_1\right) < E\left(\ln e^{\beta+b_1} + \bar{\varepsilon}_1 e^{-(\beta+b_1)}\right) \\
&= \beta + E\left(\bar{\varepsilon}_1 e^{-(\beta+b_1)}\right) = \beta + E(\bar{\varepsilon}_1) E e^{-(\beta+b_1)} = \beta.
\end{aligned}
$$

The expectation (3.9) can be calculated exactly as a two-dimensional integral or approximately based on the following general second-order approximation formula

$$
E r(u,v) = r(0,0) + \frac{1}{2}\mathrm{var}(u) \left.\frac{\partial^2 r}{\partial u^2}\right|_{u=0,v=0} + \frac{1}{2}\mathrm{var}(v) \left.\frac{\partial^2 r}{\partial v^2}\right|_{u=0,v=0}, \qquad (3.10)
$$

where $r = r(u,v)$ is any function of independent random variables u and v, and $E(u) = E(v) = 0$. To apply (3.10) to (3.9) we set $u = b_1$, $v = \bar{\varepsilon}_1$, and $r(u,v) = e^{\beta+u}+v$. Remarkably, the second derivative with respect to random effect is zero, i.e. the second term in the approximation disappears (this is a general phenomenon, see section 5). Since $\mathrm{var}(b_1) = \sigma^2\omega^2$ and $\mathrm{var}(\bar{\varepsilon}_1) = \sigma^2/n$, the asymptotic bias for fixed n can be approximated as

$$
\lim_{N\to\infty} \hat{\beta}_{TS} - \beta \simeq -\frac{\sigma^2}{2n} e^{-2\beta}. \qquad (3.11)
$$

As we see, the bias is negative and has the order of $1/n$. Also, the bias does not depend on the variance of the random effects, ω^2. In particular, if σ^2 is close to zero one can expect unbiased estimation by the two-stage method.

There is one well-known drawback of the TS-estimator: it may not exist when n is fixed and particularly small. Indeed, for the considered exponential model formula (3.8) fails when $\bar{y}_i \leq 0$. Since $\bar{y}_i$ is a continuous variable with the range $(-\infty, \infty)$ and all $\bar{y}_i$ are iid, $\Pr(\bar{y}_i > 0) = \Pr(\bar{y}_1 > 0) = q_n < 1$ and

$$
\Pr(\hat{\beta}_{TS} \text{ does not exist}) = 1 - q_n^N \to 1, \qquad N \to \infty
$$

because q_n is constant for fixed n. Therefore, the TS-estimator, as it is defined in (3.8) does not exist with probability 1 for fixed n! In fact, it may not exist even if $n \to \infty$. To obtain existence with probability 1 we must satisfy the following limit: $N \ln q_n \to 0$. Since $q_n \to 0$ when $n \to \infty$ it means that n must be large enough to satisfy the above condition.

There is a remedy: if $\bar{y}_i \leq 0$, we simply do not include the according term in the sum (3.8). Hence, the corrected TS-estimator becomes

$$
\tilde{\beta}_{TS} = \frac{\sum_{\bar{y}_i>0} \ln \bar{y}_i}{\#(y_i > 0, i = 1,...,N)} \simeq \frac{\sum_{\bar{y}_i>0} \ln \bar{y}_i}{N q_n},
$$

which exists now with probability one. The approximate formula for the bias works for $\tilde{\beta}_{TS}$ as well. Now we can prove the consistency of the corrected TS-estimator when $n \to \infty$:

$$
\begin{aligned}
&\lim_{N\to\infty} \lim_{n\to\infty} \tilde{\beta}_{TS} \\
&= \lim_{N\to\infty} \frac{1}{Nq} \sum_{\bar{y}_i>0} \ln\left(\lim_{n\to\infty} (\bar{y}_i \mid \bar{y}_i > 0\right)
\end{aligned}
$$

$$= \lim_{N\to\infty} \frac{1}{Nq} \sum_{i=1}^{N} q \ln\left(e^{\beta+b_i} + \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{ij} \,\Big|\, e^{\beta+b_i} + \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{ij} > 0 \right)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (\beta + b_i) = \beta + \lim \frac{1}{N} \sum_{i=1}^{N} b_i = \beta.$$

As follows from (3.9), the exact bias can be calculated as a two-dimensional integral

$$E(\widetilde{\beta}_{TS}) - \beta = \frac{G}{q_n} - \beta \qquad (3.12)$$

where

$$G = \frac{\sqrt{n}}{2\pi\omega\sigma^2} \iint_{e^{\beta+x}+y>0} \ln(e^{\beta+x} + y) \exp\left(-\frac{1}{2\sigma^2\omega^2} x^2 - \frac{n}{2\sigma^2} y^2 \right) dxdy;$$

and

$$q_n = \frac{\sqrt{n}}{2\pi\omega\sigma^2} \iint_{e^{\beta+x}+y>0} \exp\left(-\frac{1}{2\sigma^2\omega^2} x^2 - \frac{n}{2\sigma^2} y^2 \right) dxdy;$$

the probability of the existence of the corrected TS-estimator.

3.4. The Lindstrom-Bates estimator

Since σ^2 and ω^2 are known, we implement only the penalized nonlinear least squares step. The Lindstrom-Bates (LB)-estimator, $\widehat{\beta}_{LB}$ is the solution to the following optimization problem:

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{n} \sum_{j=1}^{n} \left(y_{ij} - e^{\beta+\tau_i} \right)^2 + \omega^{-2} \tau_i^2 \right) \Rightarrow \min_{\beta,\tau_1,\ldots,\tau_N}.$$

The estimating equations for the estimates of random effects $\tau_1, \ldots, \tau_N$ and $\widehat{\beta}_{LB}$ are:

$$e^{2(\beta+\tau_i)} - e^{\beta+\tau_i} \overline{y}_i + \frac{1}{n\omega^2} \tau_i = 0; \quad i = 1, \ldots, N; \qquad (3.13)$$

$$\frac{1}{N} \sum_{i=1}^{N} \left(e^{2(\beta+\tau_i)} - e^{\beta+\tau_i} \overline{y}_i \right) = 0. \qquad (3.14)$$

Summing the equations (3.13) over i and substituting the sum into (3.14) leads to the estimating equation for the LB-estimator:

$$\frac{1}{N} \sum_{i=1}^{N} \widehat{\tau}(\beta; \overline{y}_i) = 0 \qquad (3.15)$$

where $\widehat{\tau}(\beta; \overline{y}_i)$ is the solution to (3.13) β as a function of β and $\overline{y}_i$. The LB-estimator is also inconsistent when $n = \text{const}$. To determine the systematic bias the following fact will be used. Let $u_1, u_2, \ldots u_N$ be iid random variables with a distribution dependent on a parameter β; the true parameter is denoted by β_0. Let the estimating equation for $\widehat{\beta}_N$ be $\frac{1}{N} \sum_{i=1}^{N} S(u_i; \beta) = 0$. Then $\lim_{N\to\infty} \widehat{\beta}_N = \beta_*$ where β_* is the solution to the equation $\overline{S}(\beta; \beta_0) := E_{\beta_0} S(u_1; \beta) = 0$ (e.g., Huber 1981). Based on this fact, in order to compute the bias of $\widehat{\beta}_{LB}$ we introduce function $\overline{S}(\beta; \beta_0) = E_{\beta_0} \widehat{\tau}(\beta; \overline{y})$ where $\widehat{\tau}(\beta; \overline{y})$ is the solution to the nonlinear equation

$$e^{2(\beta+\tau)} - e^{\beta+\tau} \overline{y} + \tau/(n\omega^2) = 0. \qquad (3.16)$$

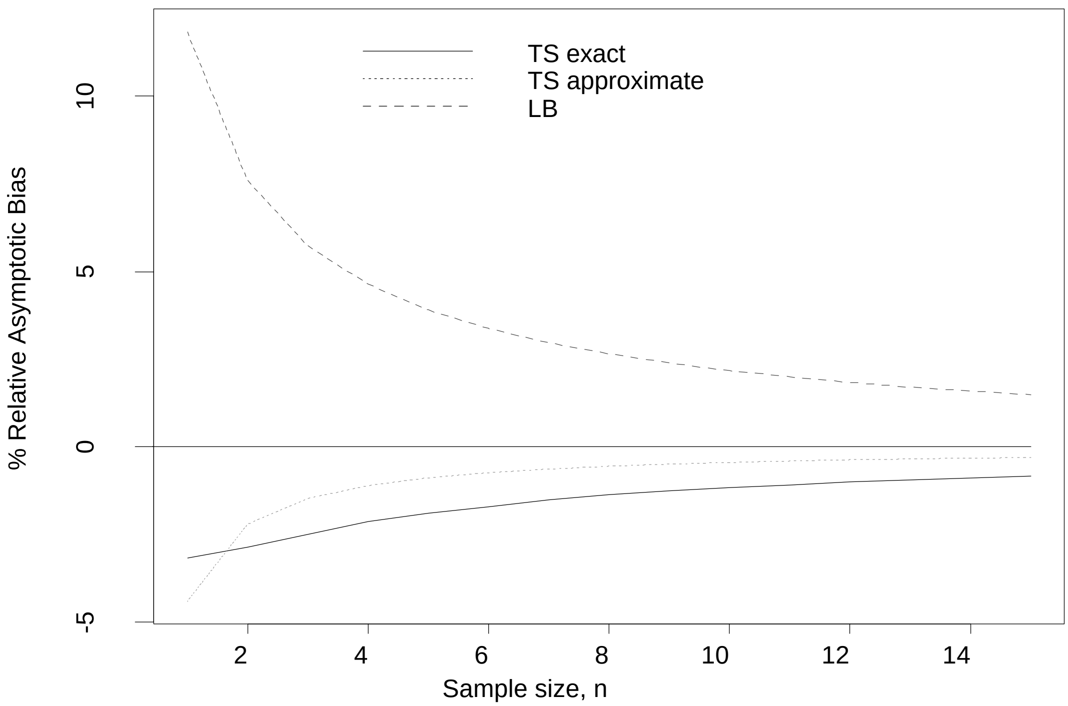

Figure 3.1: The asymptotic relative bias of three estimators for the one-parameter balanced exponential model (3.1), (3.2) when $N \to \infty$ for different n $(\beta = 0.7;\ \sigma = 0.5;\ \omega = 1.5)$.

Then the asymptotic limit of $\hat\beta_{LB}$, when $N \to \infty$ and n is fixed, is the root of the equation $\bar S(\beta; \beta_0) = 0$. Thus, the exact asymptotic limit for $\hat\beta_{LB}$ is computed based on the function

$$\bar S(\beta;\beta_0) = \frac{\sqrt{n}}{2\pi\sigma^2\omega}\int\int \hat\psi(\beta; e^{\beta_0+x}+y)\exp\left(-\frac{x^2}{2\sigma^2\omega^2}-\frac{ny^2}{2\sigma^2}\right)dxdy.$$

In Figure 3.1 we compare the relative asymptotic bias $(\hat\beta - \beta_0)/\beta_0 \cdot 100\%$ of three estimators for $\sigma = 0.1$, $\omega = 2$, $\beta_0 = 0.5$ when $N \to \infty$ and n is fixed. The VC-estimator is not shown since its bias $.5\sigma^2\omega^2$ does not depend on n, with the relative bias 40%. As we can see the approximation formula for the TS-estimator bias works quite well. The probability of nonexistence of the TS-estimator is low: for $n = 1$ we have $1 - q = 0.0182$ which drops to 0.0003 for $n = 15$. It is interesting to observe that the bias of the LB-estimator for this model is positive and larger in absolute value than of the TS-estimator.

4. The equivalence of the ML, LB and TS estimators when $N \to \infty$ and $\min n_i \to \infty$

The VC-estimator is dropped from the asymptotic consideration because it is not consistent even when $\min n_i \to \infty$, as was shown in the previous section.

Theorem 4.1. Under mild asymptotic assumptions the maximum likelihood, the two-stage and Lindstrom-Bates estimators have the same limiting distribution, i.e., are asymptotically equivalent when $N \to \infty$ and $\min n_i \to \infty$.

As follows from the proof all three estimators: the ML, LB and TS are equivalent to the following simplified TS-estimator

$$\widehat{\boldsymbol{\beta}} = \left(\sum_1^N Z_i' \widehat{\mathbf{D}}_1^{-1} Z_i\right)^{-1} \left(\sum_1^N Z_i' \widehat{\mathbf{D}}_1^{-1} \widehat{\boldsymbol{\alpha}}_i^{LS}\right) \tag{4.1}$$

where $\widehat{\mathbf{D}}_1$ is defined by (2.12). This estimator follows directly from (2.11) because $V_i \to 0$ and $\widehat{\mathbf{D}}_2 \to 0$ with probability 1 when $\min n_i \to \infty$.

To assess the quality of the ML and TS-estimators for fixed $n_i = n$, we compute the asymptotic MSE for the univariate balanced exponential model (3.1), (3.2). The LB-estimator is dropped from the analysis because it is equivalent to the TS-estimator. Recall, that the absolute lower bound for the MSE is $\sigma^2\omega^2$. The asymptotic MSE for the TS-estimator is calculated as a two-dimensional integral

$$MSE_{TS} = \frac{1}{q_n}\frac{\sqrt{n}}{2\pi\omega\sigma^2}\int\int_{e^{\beta+x}+y>0}\left(\beta - \ln(e^{\beta+x}+y)\right)^2 \exp\left(-\frac{1}{2\sigma^2\omega^2}x^2 - \frac{n}{2\sigma^2}y^2\right) dxdy,$$

where q_n is given in section 3.3. The MSE for the MLE is equal to I_β^{-1} where I_β is defined by (3.3).

In Figure 4.1 the asymptotic MSE is shown for different values of n and parameters defined in the previous section. The MSE for the MLE is less than for the TS-estimator, as one could expect from the maximum likelihood theory. When $n \to \infty$ the MSE approaches to its absolute lower bound $\sigma^2\omega^2 = 0.56$.

5. Bias-corrected two-stage estimator

As follows from the proof of the above theorem the bias of the TS-estimator is driven by the fact that for fixed n_i the individual LS-estimates $\widehat{\boldsymbol{\alpha}}_i^{LS}$ are biased (that fact was also pointed out by Vonesh and Carter 1992). Therefore, to find a bias correction to the TS-estimator we have to evaluate the conditional bias in $\widehat{\boldsymbol{\alpha}}_i^{LS}$, i.e., $E(\widehat{\boldsymbol{\alpha}}_i^{LS}|a_i) - a_i$. Omitting i the normal equation for the individual LS-estimator can be written as $R(\alpha)(y - f(\alpha)) = 0$, or in terms of ε and b as

$$R(\alpha)\left(f(Z\beta + b) + \varepsilon - f(\alpha)\right) = 0. \tag{5.1}$$

The solution to (5.1), $\widehat{\boldsymbol{\alpha}}^{LS}$ is an implicit function of ε and b. We aim to evaluate the bias using the second-order approximation to $\widehat{\boldsymbol{\alpha}}^{LS}$ as the function of b and ε in the neighborhood $\varepsilon = 0$ and $b = 0$. Therefore, one can expect that this approximation will work well especially for small σ^2 and D. This formula can be considered as a generalization of (3.11). Since ε and b are independent the cross-derivative vanishes and we only need to calculate

$$\left.\frac{\partial^2 \widehat{\boldsymbol{\alpha}}^{LS}}{\partial \varepsilon^2}\right|_{\varepsilon=0, b=0} \quad \text{and} \quad \left.\frac{\partial^2 \widehat{\boldsymbol{\alpha}}^{LS}}{\partial b^2}\right|_{\varepsilon=0, b=0}. \tag{5.2}$$

Notice that for $\varepsilon = 0$ and $b = 0$ we have $\widehat{\boldsymbol{\alpha}}^{LS} = Z\beta$.

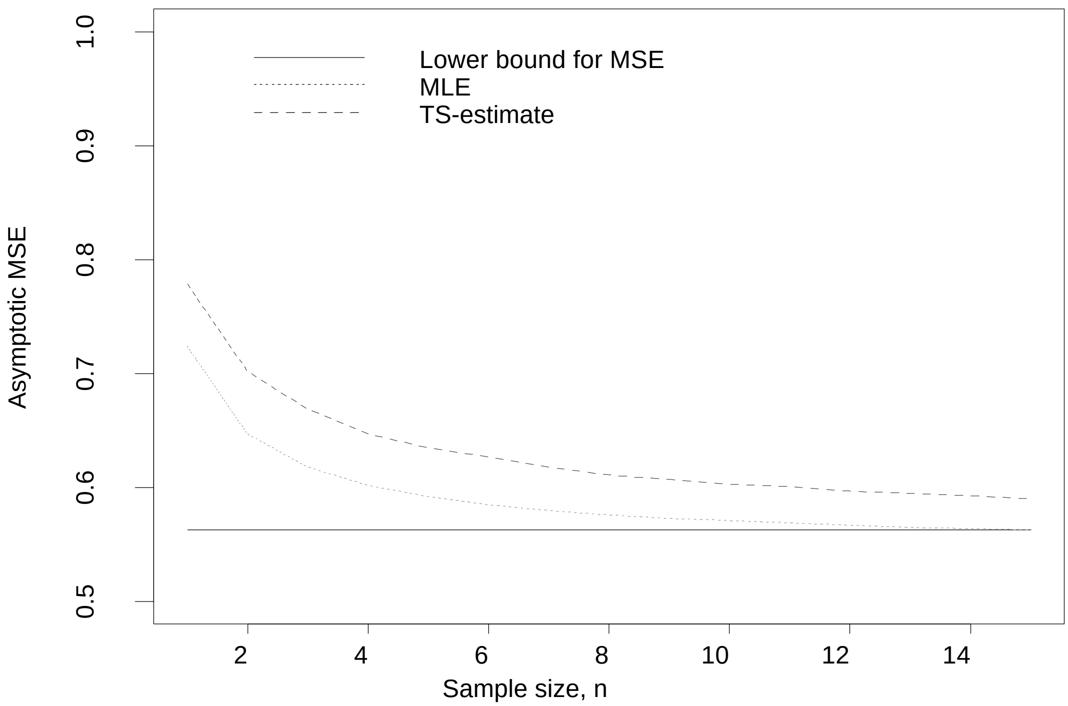

Figure 4.1: The asymptotic MSE of the MLE and two-stage estimate for the one-parameter balanced exponential model when $N \to \infty$ and n is fixed. The lower bound for the variance is $\sigma^2\omega^2 = 0.56$.

As we have learned from the exponential model the second derivative of the TS-estimator with respect to the random effect is zero. Now we shall show that it is true for any nonlinear model. First, let us consider the univariate case, i.e., α and b are scalars. Then (5.1) can be rewritten as

$$(f(\gamma + b) + \varepsilon - f(\alpha))' \dot{f}(\alpha) = 0 \tag{5.3}$$

where $\gamma = Z\beta$ and $\dot{f}$ is the $n \times 1$ vector of the first derivative. Differentiating (5.3) with respect to b we obtain

$$\left(\dot{f}(\gamma + b) - \dot{f}(\alpha)\frac{\partial \alpha}{\partial b} \right)' \dot{f}(\alpha)$$

$$+ (f(\gamma + b) + \varepsilon - f(\alpha))' \ddot{f}(\alpha)\frac{\partial \alpha}{\partial b} = 0. \tag{5.4}$$

Substituting $\varepsilon = 0$ and $b = 0$ we obtain

$$\left. \frac{\partial \hat{\alpha}_{LS}}{\partial b} \right|_{\varepsilon = 0, b = 0} = 1. \tag{5.5}$$

To find the second derivatives at $\varepsilon = 0, b = 0$ we observe that we do not need to differentiate the second factors in (5.4) because the first factors vanish at $\varepsilon = 0, b = 0$. Therefore, differentiating the first factors in (5.4) with respect to b we obtain

$$\left(\ddot{f}(\gamma+b)-\ddot{f}(\alpha)\left(\frac{\partial\alpha}{\partial b}\right)^2-\dot{f}(\alpha)\frac{\partial^2\alpha}{\partial b^2}\right)'\dot{f}(\alpha)$$

$$+\left(\dot{f}(\gamma+b)-\dot{f}(\alpha)\frac{\partial\alpha}{\partial b}\right)'\ddot{f}(\alpha)\frac{\partial\alpha}{\partial b}=0.$$

Calculating the last expression at $\varepsilon=0, b=0$ we obtain

$$\left(\ddot{f}(\gamma)-\ddot{f}(\gamma)\left(\frac{\partial\alpha}{\partial b}\right)^2-\dot{f}(\gamma)\frac{\partial^2\alpha}{\partial b^2}\right)'\dot{f}(\gamma)=0.$$

Finally, using (5.5), we come to

$$\left.\frac{\partial^2\widehat{\alpha}^{LS}}{\partial b^2}\right|_{\varepsilon=0,b=0}=0,$$

because $\|\dot{f}(\gamma)\|\neq 0.$ In multivariate case we can repeat these steps for every component of vectors b and α. Thus, we conclude that the second derivative for individual LS with respect to random effects at zero is zero.

In order to find the bias for $\widehat{\alpha}^{LS}$ driven by the error term ε we adopt the theory on the bias in nonlinear estimation based on the second-order approximation developed by Box (1971). The formula for the bias he derived is

$$\widehat{\alpha}^{LS}-E\widehat{\alpha}^{LS}\simeq-\frac{\sigma^2}{2}(R'R)^{-1}R\mathbf{d} \tag{5.6}$$

where $\mathbf{d}$ is the $n\times 1$ vector with the jth element

$$d_j=\mathrm{tr}\left((R'R)^{-1}H_j\right);\quad \text{where}\quad H_j=\frac{\partial^2 f_j}{\partial\alpha^2};\quad j=1,...,n.$$

Since the effect of random effects on the bias is zero formula for bias (5.6) should approximate the overall effect of random terms ε and $\mathbf{b}$ well at least when variances of ε and $\mathbf{b}$ are small. Finally, the bias-corrected version of the TS-estimator has the form

$$\widehat{\beta}^c_{TS}=\left[\sum Z_i'(V_i+\widehat{D})^{-1}Z_i\right]^{-1}\left[\sum Z_i'(V_i+\widehat{D})^{-1}\left(\widehat{\alpha}_i^{LS}+\mathbf{c}_i\right)\right]$$

where

$$\mathbf{c}_i=\frac{1}{2}\widehat{\sigma}^2(R_i'R_i)^{-1}R_i\mathbf{d}_i;\quad H_{ij}=\frac{\partial^2 f_{ij}}{\partial\alpha^2}$$

and the jth element of $n_i\times 1$ vector $\mathbf{d}_i$ is $\mathrm{tr}\left((R_i'R_i)^{-1}H_{ij}\right).$

6. Discussion

Methods of estimation for nonlinear mixed-effects models based on a straightforward linear approximation are not consistent regardless of the number of observations per individual. This may be explained in the following way. Let $E(y|a)=f(a)$ be a simple statistical model with a random variable $a \sim N(\beta; \tau^2)$ inside a nonlinear function f: We aim to extract an intrinsic nonlinear random variable a via the first-order approximation substituting $f(a)$ by $f(\beta) + (a - \beta) f'(\beta)$. However, these two models have different means. For instance, assuming that f is convex, i.e. $f'' > 0$ we have

$$
\begin{aligned}
Ef(a) &= E\left(f(\beta) + (a - \beta) f'(\beta) + .5(a - \beta)^2 f''(\xi)\right) \\
&> E\left[f(\beta) + (a - \beta) f'(\beta)\right] = f(\beta)
\end{aligned}
$$

with a positive difference of order $.5\tau^2 f''(\beta)$: Since the approximate model has a systematic unremovable bias it is not surprising that all estimators based on this model are inconsistent. Similar arguments can be used to show that higher order approximations encounter the same phenomenon (e.g., Solomon and Cox 1992). Therefore, applying any Taylor series expansion one can only reduce the bias, not eliminate it completely when $N \to \infty$ and $\{n_i\}$ are bounded.

Furthermore, neither the two-stage, nor the Lindstrom-Bates estimator are consistent when the number of observations per individual is finite. Therefore, is there a reason to use a cumbersome procedure if it asymptotically equivalent to a much easier two-stage procedure? The TS-procedure was criticized (e.g., Vonesh and Carter 1992) for the fact that it requires a large number of observations per individual. However, for NLME model, the same requirement is needed for LB-procedure. Comparisons of the TS and VC estimators to the LB -estimator based on statistical simulations in real nonlinear models confirm basic conclusions of the asymptotic analysis of the present paper (Stukel and Demidenko 1996). Further work should be done to investigate statistical properties of the bias-corrected TS-estimator suggested in this paper.

On contrary, the maximum likelihood method does not require a large number of observations per individual. Therefore, in practice we would recommend to use it, especially when n_i are relatively small. The number of random effects is a critical point for the integration problem. In particularly, when there is only one random effect the integration problem is not serious. Pinheiro and Bates (1995) came to a conclusion that the Lindstrom-Bates estimate might be a good starting point for the maximum likelihood procedure. Keeping in mind the increased power of computers we cannot excuse ourselves avoiding "computational intensive" methods: in most cases costs of collecting data exceeds costs of developing, programming, and running 'computational intensive' procedures.

The asymptotic issues are of particular importance when experiments are designed. Given the ability to increase the number of individuals and the number of observations per individual, what way to go, what is more important N or n_i? Hopefully, this study can shed some light on this question. However, one should remember that applying asymptotic properties to finite sample sizes could be misleading.

ACKNOWLEDGMENTS

I thank Therese Stukel, Edward Vonesh and Jose Pinheiro for helpful comments. This work was supported by National Cancer Institute grants CA52192 and CA50597.

References

[1] Berkey, C.S. and Laird, N.M. (1986). Nonlinear growth curve analyses: estimating population parameters. Annals of Human Biology 13, 111-128.

[2] Box, M.J. (1971). Bias in nonlinear estimation. Journal of the Royal Statistical Society, ser. B, 33, 171-190.

[3] Davidian, M. and Gallant, A.R. (1992). Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. Journal of Pharmacokinetics and Biopharmaceutics 20, 529-556.

[4] Davidian, M. and Giltinan, D.M. (1995). Nonlinear Models for Repeated Measurement Data. London: Chapman and Hall.

[5] Demidenko, E. (1996). On the existence of the least squares estimate in nonlinear growth curve models of exponential type. Communications in Statistics, ser. Theory and Methods 25, 159-182.

[6] Gallant, A.R. (1987). Nonlinear Statistical Models. New York: John Wiley.

[7] Gourieroux, C, Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: theory. Econometrica 52, 681-700.

[8] Graybill, F.A. (1983). Matrices with Applications in Statistics, Belmont: Wadsworth Inc.

[9] Huber, P.J. (1981). Robust Statistics. New York: Wiley.

[10] Laird, N.M. and Ware, J.H. (1982). Random-e¤ects models for longitudinal data. Biometrics 38, 963-974.

[11] Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed-e¤ects models for repeated measures. Biometrics 46, 673-687.

[12] Neuhaus, J.M. and Lesperance, M.L. (1996). Estimation e¢ciency in a binary mixed-e¤ects model setting. Biometrika 83, 441-446.

[13] Parke, W.R. (1986). Pseudo maximum likelihood estimation: the asymptotic distribution. Annals of Statistics 14, 355-357.

[14] Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-e¤ects model. Journal of Computational and Graphical Statistics 4, 12-35.

[15] Pocock, S.J., Cook, D.G. and Beresford, S.A.A.. (1981). Regression of area mortality rates on explanatory variables: what weighting is appropriate? Applied Statistics 30, 286-295.

[16] Rao, C.R.(1973). Linear Statistical Inference and its Applications. Second Edition. New York: Wiley.

[17] Schervish, M.J. (1995). Theory of Statistics. New York: Springer-Verlag.

[18] Sheiner, L.B. and Beal, S.L. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. Journal of Pharmacokinetics and Biopharmaceutics 8, 553-571.

[19] Sheiner, L.B. and Beal, S.L. (1985). Pharmacokinetic parameter estimates from several least squares procedures: superiority of extended least squares. Journal of Pharmacokinetics and Biopharmaceutics 13, 185-201.

[20] Solomon, P.J. and Cox, D.R. (1992). Nonlinear component of variance models. Biometrika 79, 1-11.

[21] Stiratelli, R., Laird, N.M. and Ware, J.H. (1984). Random-e¤ects models for serial observations with binary response. Biometrics 40, 961-971.

[22] Stukel, T. and Demidenko, E. (1997). Two-stage estimation for general linear growth curve models. Biometrics 53, 340-348.

[23] Stukel, T. and Demidenko, E. (1997). Comparisons of methods of estimation for nonlinear mixed-e¤ects models, in : "Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions". Springer Lectures in Statistics. New York: Springer-Verlag.

[24] Vonesh, E.F. and Carter R.L. (1987). E¢cient inference for random-coe¢cient growth curve models with unbalanced data. Biometrics 43, 617-628.

[25] Vonesh, E.F. and Carter R.L. (1992). Mixed-e¤ects nonlinear regression for unbalanced repeated measures. Biometrics 48, 1-17.

[26] Vonesh, E.F. (1992). Non-linear models for the analysis of longitudinal data. Statistics in Medicine 11, 1929-1954.

[27] Vonesh, E.F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-e¤ects models. Biometrika 83, 447-452.

[28] Vonesh, E.F. and Chinchilli, V.M. (1997). Linear and Nonlinear Models for the Analysis of Repeated Measurements. New York: Marcel Dekker.

[29] Wol…nger, R. (1993). Laplace's approximation for nonlinear mixed-e¤ects models. Biometrika 80, 791-795.

7. Appendix. Proofs

7.1. Lemma 2.1

We use the following vector analog of the Cauchy inequality (e.g., Rao 1973): if $d\mu(x)$ is a measure and $u(x)$ is an integrable vector function and $f(x)$ is an integrable scalar function then

$$\left(\int u(x)f(x)d\mu(x)\right)\left(\int u(x)f(x)d\mu(x)\right)' \tag{7.1}$$

$$\le \left(\int f^2(x)d\mu(x)\right)\left(\int u(x)u'(x)d\mu(x)\right). \tag{7.2}$$

Omitting, for simplicity of notation, arguments of functions we obtain

$$\frac{\partial \ln f}{\partial\theta} = \left(\int f_1 f_2 da\right)^{-1}\left(\int f_1 \frac{\partial f_2}{\partial\theta} da\right)$$

and

$$I_Y = \int \left\{\left(\int f_1 \frac{\partial f_2}{\partial\theta} da\right)\left(\int f_1 \frac{\partial f_2}{\partial\theta} da\right)'\left(\int f_1 f_2 da\right)^{-1}\right\} dy. \tag{7.3}$$

But by the inequality (7.1),

$$\begin{aligned}
&\left(\int f_1 \frac{\partial f_2}{\partial\theta} da\right)\left(\int f_1 \frac{\partial f_2}{\partial\theta} da\right)' \\
&= \left[\int \left(\sqrt{\frac{f_1}{f_2}}\frac{\partial f_2}{\partial\theta}\right)\left(\sqrt{f_1 f_2}\right) da\right]\left[\int \left(\sqrt{\frac{f_1}{f_2}}\frac{\partial f_2}{\partial\theta}\right)\left(\sqrt{f_1 f_2}\right) da\right]' \\
&\le \left[\int \left(\frac{\partial f_2}{\partial\theta}\right)\left(\frac{\partial f_2}{\partial\theta}\right)' \frac{f_1}{f_2} da\right]\left(\int f_1 f_2 da\right)
\end{aligned} \tag{7.4}$$

Substituting it into (7.3) we come to

$$\begin{aligned}
I_Y &\le \int \left[\int \left(\frac{\partial f_2}{\partial\theta}\right)\left(\frac{\partial f_2}{\partial\theta}\right)' \frac{f_1}{f_2} da\right] dy \\
&= \int \left[\frac{1}{f_2}\left(\frac{\partial f_2}{\partial\theta}\right)\left(\frac{\partial f_2}{\partial\theta}\right)'\left(\int f_1 dy\right)\right] da = I_Y
\end{aligned}$$

because $\int f_1 dy = 1$ for any a.

7.2. Theorem 4.1

The plan of the proof is as follows:

1. The LB and the TS -estimators are equivalent when $\min n_i \to \infty$ and σ^2, β are fixed.
2. The TS-estimator is consistent and normally distributed with the covariance matrix equal to the lower bound (2.7).
3. The LB and TS-estimators produce consistent estimators for σ^2 and β.
4. The TS-estimator is equivalent to the MLE when $N \to \infty; \min n_i \to \infty$ and β is substituted by a consistent estimator.

The 'mild' asymptotic assumptions are: (i) the probability that the individual LS-estimator for (2.9) exists is fixed and less than 1, (ii) elements of matrices $\{Z_i\}$ are bounded and the limit matrix $\lim N^{-1}\sum Z_i'\Omega^{-1}Z_i = M_z$ is nonsingular (iii) standard asymptotic assumptions for the ith nonlinear regression of (2.1), conditionally on a_i, are fulfilled (e.g. Gallant, 1989). The latter implies that the individual nonlinear least squares estimators $\hat{\alpha}_i^{LS}$ are consistent for a_i, asymptotically normally distributed and efficient when $n_i \to \infty$.

1. Before proving the equivalence between LB and TS-estimators we will prove the following simple result of optimization theory. Let $\hat{\beta}(\delta)$ be the solution to the following minimization problem

$$\sum_1^N \left(\|y_i - f_i(Z_i\beta + b_i)\|^2 + \delta b_i'\Omega^{-1}b_i \right), \qquad \delta > 0. \tag{7.5}$$

Then

$$\lim_{\delta \to 0} \hat{\beta}(\delta) = \hat{\beta}, \tag{7.6}$$

where $\hat{\beta}$ is defined as

$$\hat{\beta} = \left(\sum_1^N Z_i'\Omega^{-1}Z_i\right)^{-1}\left(\sum_1^N Z_i'\Omega^{-1}\hat{\alpha}_i^{LS}\right), \tag{7.7}$$

and $\hat{\alpha}_i^{LS}$ minimizes the individual sum of squares, $\|y_i - f_i(\alpha_i)\|^2$. To prove this, let us rewrite the function (7.5) as

$$\sum \left(\|y_i - f_i(\alpha_i)\|^2 + \delta(\alpha_i - Z_i\beta)'\Omega^{-1}(\alpha_i - Z_i\beta) \right).$$

The first order conditions for the optimization problem are:

$$\frac{\partial f_i}{\partial \alpha_i}(y_i - f_i(\alpha_i)) + 2\delta\Omega^{-1}(\alpha_i - Z_i) = 0, \quad i = 1, \ldots, N \tag{7.8}$$

and

$$Z_i'\Omega^{-1}(\alpha_i - Z_i\beta) = 0. \tag{7.9}$$

Since $\delta \to 0$ the equations (7.8) become equivalent to $\frac{\partial f_i}{\partial \alpha_i}(y_i - f_i(\alpha_i)) = 0$ that lead to the individual least squares solution, $\alpha_i = \hat{\alpha}_i^{LS}$. Substituting them into (7.9) we come to (7.7), the limit (7.6) is proved.

Now we prove that the LB-estimator is equivalent to (7.7) asymptotically with probability 1. Since Ω is fixed in the LB procedure we have to consider only the penalized nonlinear least squares step:

$$\sum_1^N \left(\frac{1}{n_i}\|y_i - f_i(Z_i\beta + b_i)\|^2 + \frac{1}{n_i}b_i'\Omega^{-1}b_i \right) \Rightarrow \min_{\beta, b_1, \ldots, b_N}. \tag{7.10}$$

To apply (7.6) to (7.10) we show that there exists a constant A that with the probability 1

$$\frac{1}{n_i}\|y_i - f_i(Z_i\beta + b_i)\|^2 \geq A > 0, \quad i = 1, \ldots, N \to \infty. \tag{7.11}$$

when $\min n_i \to \infty$: In fact, letting $p_i = Z_i\beta + b_i$; conditionally on random variables a_i; we have

$$\begin{aligned}
&\lim_{n_i\to\infty}\frac{1}{n_i}\| y_i - f_i(p_i)\|^2 \\
&\quad = \lim_{n_i\to\infty}\frac{1}{n_i}\| y_i - f_i(a_i)\|^2 \\
&\quad\quad + \lim_{n_i\to\infty}\frac{2}{n_i}(y_i - f_i(a_i))'(f_i(a_i) - f_i(p_i)) \\
&+ \lim_{n_i\to\infty}\frac{1}{n_i}\| f_i(a_i) - f_i(p_i)\|^2 .
\end{aligned}$$

However, by standard arguments of nonlinear regression theory, conditionally on a_i; with probability 1 uniformly for $i = 1,\dots,N$ we have

$$\lim_{n_i\to\infty}\frac{1}{n_i}\| y_i - f_i(a_i)\|^2 = \sigma^2;\quad \lim_{n_i\to\infty}\frac{2}{n_i}(y_i - f_i(a_i))'(f_i(a_i) - f_i(p_i)) = 0;$$
$$\lim_{n_i\to\infty}\frac{1}{n_i}\| f_i(a_i) - f_i(p_i)\|^2 \geq 0.$$

Therefore in (7.11) we can set $A = \sigma^2$. Denoting $\delta = 1/n_i$ and applying the above optimization result we prove that the LB estimator is equivalent to (7.7) with probability 1 when $\min n_i \to \infty$.

Now we show that the TS-estimator (2.11) is equivalent to (7.7). As was mentioned above, from the standard theory of nonlinear regression, conditionally on a_i we have

$$\lim_{n_i\to\infty}\hat\alpha_i^{LS} = a_i;\qquad \sqrt{n_i}\left(\hat\alpha_i^{LS} - a_i\right) \sim N\left(0, \sigma^2 V_i^{-1}\right) \tag{7.12}$$

where

$$V_i = E_{x_i}\left[R_i(x_i;a_i)R_i'(x_i;a_i)\,\middle|\, a_i\right];\quad R_i(x_i;a_i) = \left.\frac{\partial f(a_i;x_i)}{\partial\alpha_i}\right|_{\alpha_i=a_i}.$$

As follows, $V_i = O(n_i^{-1})$ and $\Omega_2 \to 0$ Thus for given σ^2 and Ω; the TS-estimator (2.11) asymptotically is equivalent to (7.7).

2. Starting from this point we can deal only with the second stage model (2.2) assuming that instead of observations a_i we substitute $\hat\alpha_i^{LS}$ with the properties (7.12). Let us denote $\delta_i = a_i - \hat\alpha_i^{LS}$; then $\sqrt{N}(\hat\beta - \beta) = s_N + u_N$ where

$$\begin{aligned}
s_N &= \left(\frac{1}{N}\sum_1^N Z_i'\Omega^{-1}Z_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_1^N Z_i'\Omega^{-1}b_i\right);\\
u_N &= \left(\frac{1}{N}\sum_1^N Z_i'\Omega^{-1}Z_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_1^N Z_i'\Omega^{-1}\delta_i\right).
\end{aligned}$$

Obviously, $s_N \sim N\left(0, \sigma^2 M_Z^{-1}\right)$; and since $\mathrm{cov}(\delta_i) = O(1/n_i)$ we have $u_N \to 0$ with probability 1. That proves that the estimator (7.7) asymptotically is normally distributed as $\sqrt{N}\left(\hat\beta - \beta\right) \sim N\left(0, \sigma^2 M_Z^{-1}\right)$.

3. First we prove that in (2.12) $\hat\Omega \to \Omega$. As it was mentioned above, $\hat\Omega_2 \to 0$, so we have to show that $\hat\Omega_1 \to \Omega$. We have

$$\hat\Omega_1 = \hat\Omega_{11} + \hat\Omega_{12} + \hat\Omega_{12}' + \hat\Omega_{22} \tag{7.13}$$

where

$$\widehat{\Omega}_{11} = \frac{1}{N}\sum_i \left(\widehat{\alpha}_i^{LS} - Z_i\beta\right)\left(\widehat{\alpha}_i^{LS} - Z_i\beta\right)';$$

$$\widehat{\Omega}_{12} = \frac{1}{N}\sum_i \left(\widehat{\alpha}_i^{LS} - Z_i\beta\right)\left[\beta - (\sum Z_j'Z_j)^{-1}(\sum Z_j'\widehat{\alpha}_j^{LS})\right]' Z_i';$$

$$\widehat{\Omega}_{22} = \frac{1}{N}\sum_i Z_i\left[\beta - (\sum Z_j'Z_j)^{-1}(\sum Z_j'\widehat{\alpha}_j^{LS})\right]\left[\beta - (\sum Z_j'Z_j)^{-1}(\sum Z_j'\widehat{\alpha}_j^{LS})\right]' Z_i';$$

$i,j = 1,\ldots,N.$ As follows from (7.12), when calculating asymptotic limits of these matrices, we can substitute $\widehat{\alpha}_i^{LS}$ by a_i that immediately implies $\operatorname{plim}\widehat{\Omega}_{11} = \Omega.$ Now we show that all other components in (7.13) converge to zero in probability. Indeed, $\widehat{\Omega}_{12}$ is equivalent to

$$-\frac{1}{N}\sum_i b_ib_i'Z_i(\sum Z_j'Z_j)^{-1}Z_i' - \frac{1}{N}\sum_i b_i(\sum_{j\neq i} b_j'Z_j)(\sum Z_j'Z_j)^{-1}Z_i'. \qquad (7.14)$$

Further,
$$E\left(\frac{1}{N}\sum_i b_ib_i'Z_i(\sum Z_j'Z_j)^{-1}Z_i'\right) = \frac{1}{N}\Omega\sum_i Z_i(\sum Z_j'Z_j)^{-1}Z_i' \to 0$$

because
$$\operatorname{tr}\left(\sum_i Z_i(\sum Z_j'Z_j)^{-1}Z_i'\right) = \sum_i \operatorname{tr}\left[Z_i(\sum Z_j'Z_j)^{-1}Z_i'\right] = \operatorname{tr}\left[\sum (Z_j'Z_j)^{-1}\sum(Z_i'Z_i)\right] = k. \qquad (7.15)$$

The second term in (7.14) has a zero expectation, because b_i and b_j are independent for $i \neq j.$ Let c be any unit vector, $c'c = 1.$ We will prove that

$$\operatorname{var}\left(\frac{1}{N}c'\sum_i b_ib_i'Z_i(\sum Z_j'Z_j)^{-1}Z_i'c\right) \to 0.$$

The following fact is used (e.g., Graybill, 1983): if $u \sim N(0,D)$ then $\operatorname{var}(u'Au) = 2\operatorname{tr}(AD)^2.$ Applying this result we obtain

$$\operatorname{var}\left(\frac{1}{N}c'\sum_i b_ib_i'Z_i(\sum Z_j'Z_j)^{-1}Z_i'c\right) = \frac{2\sigma^4}{N^2}\sum_i\left[c'\Omega Z_i(\sum Z_j'Z_j)^{-1}Z_i'c\right]^2$$
$$\leq \frac{2\sigma^4}{N^2}\left(c'\Omega\sum_i Z_i(\sum Z_j'Z_j)^{-1}Z_i'c\right)^2 \to 0$$

as follows from (7.15). Using similar arguments we can show that the second term in (7.14) and $\widehat{\Omega}_{22}$ vanish when $N \to \infty, \min n_i \to \infty.$ It proves that $\widehat{\Omega}$ is a consistent estimator for Ω.

Now we will show that the Lindstrom and Bates procedure, based on linear mixed-effects, provides a consistent estimator of Ω, too. As follows from 1, $Z_i\widehat{\beta} + \widehat{b}_i \sim \widehat{\alpha}_i^{LS} \sim a_i$; where $\sim$ means the asymptotic equivalence with probability 1. To prove that the linear mixed-effects model provides a consistent estimator for σ^2 and Ω based on (2.13), it suffices to show that $E(w_i) = R_iZ_i\beta$

and $\mathrm{cov}(w_i) = \sigma^2(I + R_i \Omega R_i')$ when $Z_i\hat{\beta} + \hat{b}_i = a_i$, $\hat{\beta} = \beta$. Indeed, for linear mixed-effects model $w_i = y_i - f_i(a_i) + R_i(a_i - Z_i\beta_i) + R_i Z_i\beta$, and

$$
\begin{aligned}
E(w_i) &= E\left(E(y_i - f_i(a_i) + R_i(a_i - Z_i\beta_i) + R_i Z_i\beta | a_i)\right) = R_i Z_i\beta, \\
\mathrm{cov}(w_i) &= E\mathrm{cov}(w_i|a_i) + \mathrm{cov}(E(w_i|a_i)) \\
&= E\mathrm{cov}(w_i|a_i) = \sigma^2(I + R_i \Omega R_i'),
\end{aligned}
$$

as was aimed to prove.

4. As was mentioned, in the situation $N \to \infty, \min n_i \to \infty$ the nonlinear mixed-effects model is asymptotically equivalent to the linear model $a_i \sim N(Z_i\beta, \sigma^2\Omega)$. It is well known that the information matrix for β and (σ^2, Ω) is block-diagonal. From a results of Gourieroux et al. (1984) and Parke (1986) it follows that for a block-diagonal information matrix the pseudo MLE is equivalent to the full MLE, or in other words, one obtains an asymptotically efficient estimator for β substituting a consistent estimator for Ω in the generalized least squares estimator. It implies that LB and TS-estimators are equivalent to (7.7) when Ω is substituted by its consistent estimator, or more precisely

$$
\sqrt{N}(\hat{\beta}^{LB} - \beta) \sim N\left(0, \sigma^2; M_z^{-1}\right), \quad \sqrt{N}(\hat{\beta}^{TS} - \beta) \sim N\left(0, \sigma^2 M_z^{-1}\right).
$$

It means that the covariances of the LB and TS-estimators achieve the lower bound (2.7). Therefore, all the estimators are asymptotically equivalent, the theorem is proved.