ELSEVIER Computational Statistics & Data Analysis 23 (1997) 541–556

COMPUTATIONAL STATISTICS & DATA ANALYSIS

The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data

Geert Verbeke *, Emmanuel Lesaffre

Biostatistical Centre for Clinical Trials, U.Z. St.-Rafael, Kapucijnenvoer 35, B-3000 Leuven, Belgium

Received May 1995; revised March 1996

Abstract

Maximum likelihood estimators for fixed effects and variance components in linear mixed models, obtained under the assumption of normally distributed random effects, are shown to be consistent and asymptotically normally distributed, even when the random-effects distribution is not normal. However, a sandwich-type correction to the inverse Fisher information matrix is then needed in order to get the correct asymptotic covariance matrix. Extensive simulations show that the so-obtained corrected standard errors are clearly superior to the naive uncorrected ones, especially for the parameters in the random-effects covariance matrix, even in moderate samples.

Keywords: Fixed effects; Random effects; Longitudinal models; Misspecification; Variance components

1. Introduction

A popular model for the analysis of continuous longitudinal data is the linear mixed-effects model. This can be seen as a two-stage model, where unobserved parameters (random effects) are first sampled from a 'prior' distribution and observed data are then sampled from independent normal distributions with linear functions of these random effects and of additional population parameters (fixed effects) as means. Once a parametric family has been chosen for the prior distribution, inferences about the fixed effects as well as about the prior distribution become available.

* Corresponding author.

PII S0167-9473(96)00047-3

To facilitate the computations, one usually assumes that the prior distribution is multivariate normal, with zero mean and a general covariance matrix. Since this assumption is difficult to check (see Verbeke and Lesaffre, to appear), it is of great importance to investigate its influence on inferences about the model parameters. Using simulations and the analysis of a real dataset, Butler and Louis (1992) have recently shown that wrongly specifying the prior distribution of univariate random effects has little effect on the fixed effects estimates. In Verbeke and Lesaffre (to appear), two datasets were analysed with a linear mixed-effects model with random effects drawn from a mixture of g multivariate normal distributions. It turned out that varying g had practically no effect on the fixed effects estimates, but also that the residual variance and the covariance matrix of the random effects were very stable. This suggested that the assumption about the random-effects distribution hardly influences any of the parameter estimates.

In this paper, we will first discuss some general asymptotic results of the maximum likelihood estimators of fixed effects and variance components, obtained under the assumption of normally distributed random effects. We will hereby especially emphasize the fact that the inverse Fisher information matrix only yields valid standard errors when the normality assumption for the random effects is satisfied. In case this assumption is violated, valid inferences follow from replacing the 'naive' estimate for the asymptotic covariance matrix by a so-called sandwich estimate.

Further, we will use extensive simulations to investigate the finite-sample properties of the MLEs and to compare the uncorrected naive standard errors with the corrected ones obtained from the sandwich estimator. This will be done in the context of longitudinal data, for linear mixed models with bivariate random effects drawn from 5 different parent distributions.

In Section 2, we will introduce the general linear mixed-effects model, and we will hereby explicitly distinguish the correct model from the model used for parameter estimation. In Section 3 we will discuss some general asymptotic properties of the maximum likelihood estimators for all population parameters in the model. Further, it will be shown in Section 4 how the asymptotic standard errors obtained from classical likelihood theory need to be corrected for possible non-normality of the random effects. Finally the performance, in finite samples, of both the uncorrected and corrected standard errors will be investigated and compared in an extensive simulation study, which we will report in Section 5.

2. Model specification

Let us first describe the correct model. Suppose $\boldsymbol{y}_i = X_i\boldsymbol{\alpha}^* + Z_i\boldsymbol{b}_i + \boldsymbol{e}_i$ is the $n_i \times 1$ vector of responses for the ith individual, $i = 1,\ldots,N$, X_i and Z_i are $n_i \times p$ and $n_i \times q$ covariate matrices of rank p and q respectively. The vector $\boldsymbol{\alpha}^*$ is p-dimensional and describes the population mean. The vector $\boldsymbol{b}_i$ is q-dimensional and represents the subject-specific regression coefficients, it has zero mean and density function

$f^*(\boldsymbol{b}_i|\boldsymbol{\psi}^*)$ with $\boldsymbol{\psi}^*$ a vector of unknown parameters. The error terms $\boldsymbol{e}_i$ are assumed to be normally distributed with zero mean and covariance matrix $\sigma^{*2}I_{n_i}$. Further, they are assumed to be independent of the random effects $\boldsymbol{b}_i$. As a result, marginally, $\boldsymbol{y}_i$ has a mean equal to $X_i\boldsymbol{\alpha}^*$ and a covariance matrix $Var(\boldsymbol{y}_i) = Z_iD^*Z_i' + \sigma^{*2}I_{n_i} = V_i^*$, with $D^* = D^*(\boldsymbol{\psi}^*) = Var(\boldsymbol{b}_i)$.

For parameter estimation it is assumed that the responses $\boldsymbol{y}_i$ can be modelled as $\boldsymbol{y}_i = X_i\boldsymbol{\alpha} + Z_i\boldsymbol{b}_i + \boldsymbol{e}_i$, but with the random effects $\boldsymbol{b}_i$ sampled from a multivariate normal distribution with zero mean and a general covariance matrix D. This is the assumption handled by the SAS (1992) procedure MIXED and the BMDP5V (1990) program. The marginal distribution for the responses $\boldsymbol{y}_i$ then becomes normal with mean $X_i\alpha$ and covariance $Z_iDZ_i' + \sigma^2 I_{n_i} = V_i$. Let $\boldsymbol{\theta}$ be the vector of all parameters in this marginal model and let Θ be the corresponding parameterspace. We then denote the corresponding distribution function by $f_i(\boldsymbol{y}_i|\boldsymbol{\theta})$. Estimates $\widehat{\boldsymbol{\alpha}}_N$, $\widehat{\sigma}_N^2$ and $\widehat{D}_N$ can now be obtained by maximizing the likelihood function $(1/N)\sum_{i=1}^N \ln f_i(\boldsymbol{y}_i|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Note that we have omitted here the superscripts $*$ to distinguish the parameters in the model used for parameter estimation from the parameters in the correct model. From now on, all expectations and variances are calculated under the correct model. Under both models, the covariance matrix of $\boldsymbol{y}_i$ is of the form $Z_i\,Var(\boldsymbol{b}_i)Z_i' + Var(e_{ij})I_{n_i}$. If $Z_i = Z$ and $n_i = q$ for all i, then the variance components $Var(\boldsymbol{b}_i)$ and $Var(e_{ij})$ are no longer identifiable. We will therefore assume from now on that $q < n_i$ for some i.

As an example, we analysed growth curves, reported by Goldstein (1979) (Table 4.3, p. 101). For 20 pre-adolescent girls, height was measured on a yearly basis from age 6 to 10. The girls were classified according to the height of their mother, which was discretized in three groups: small mothers (group A), medium mothers (group B) and tall mothers (group C). We analysed these data with a linear mixed-effects model with random intercepts and slopes, and with group, time and group×time as fixed effects:

$$\begin{aligned} &Height_{ij} \\ &\quad = b_{1i} + \alpha_1\, GroupA_i + \alpha_2\, GroupB_i + \alpha_3\, GroupC_i \\ &\qquad + \{b_{2i} + \alpha_4\, GroupA_i + \alpha_5\, GroupB_i + \alpha_6\, GroupC_i\} \times Age_{ij} + \varepsilon_{ij}. \end{aligned} \tag{1}$$

$Height_{ij}$ and Age_{ij} are the height and age respectively of the ith girl at the jth measurement. $GroupA_i$, $GroupB_i$ and $GroupC_i$ are dummy variables indicating whether or not subject i belongs to group A, B or C respectively. So, α_1, α_2 and α_3 represent the average intercepts and α_4, α_5 and α_6 the average slopes in the groups A, B and C respectively. The terms b_{1i} and b_{2i} are the random intercepts and random slopes respectively. The maximum likelihood estimates of all parameters in model (1) are given in Table 1. In Section 5, we will report the results of a simulation study where datasets were sampled from model (1) in which the parameters are replaced by the values in Table 1.

Table 1
Maximum likelihood estimates (MLE) for all fixed effects and all variance components in model (1)

	Parameter	MLE
Fixed effects		
intercept group A	α_1	81.30
intercept group B	α_2	82.97
intercept group C	α_3	83.12
slope group A	α_4	5.27
slope group B	α_5	5.57
slope group C	α_6	6.25
Covariance of random effects		
$Var(\boldsymbol{b}_i)$	D	$\begin{pmatrix} 5.99 & -0.32 \\ -0.32 & 0.11 \end{pmatrix}$
Residual variance		
$Var(\varepsilon_{ij})$	σ^2	0.48

3. Asymptotics

It follows from classical likelihood theory that, when the assumed model is correct, the MLEs are consistent and asymptotically normally distributed with the inverse Fisher information matrix as asymptotic covariance matrix. In this section and in the next section, it will be shown that these results can be extended to cases where the random effects are not normally distributed. The results hold under fairly general regularity conditions on the parameterspace Θ and on the covariates X_i and Z_i. The conditions also involve the unknown correct random-effects distribution, but they can easily be shown to be fulfilled for many distributions such as Normal, Lognormal, Gamma (including Chi-squared), Exponential, Weibull, any discrete distribution with finite support and any finite mixture of normal distributions. All technical details, including details of regularity conditions and proofs of the theorems, are deferred to Verbeke and Lesaffre (1994).

Further, we will assume D to be unstructured, i.e. we maximize the marginal likelihood with respect to $\boldsymbol{\alpha}$, σ^2 and D without assuming any specific structure for D. The SAS (1992) procedure MIXED for example, allows the maximization under several covariance structures such as compound symmetry, simple, autoregressive(1), Toeplitz, but also under some spatial covariance structures (exponential, Gaussian, ...). However, it follows from the proofs in Verbeke and Lesaffre (1994) that the results still apply if a structure for D is assumed which encompasses D^*.

Theorem 1. *Under general regularity conditions, we have that $\widehat{\boldsymbol{\alpha}}_N$, $\widehat{\sigma}^2_N$ and $\widehat{D}_N$ are strongly consistent estimators for $\boldsymbol{\alpha}^*$, σ^{*2} and D^*, when $N \to \infty$.*

The importance of the above theorem is that it assures that maximum likelihood estimation, based on the assumption of normality for the random effects $\boldsymbol{b}_i$, yields consistent estimators for all parameters in the model, including the random-effects covariance matrix, even when the true distribution of the $\boldsymbol{b}_i$ is not normal. Hence, the mean vector $X_i\boldsymbol{\alpha}^*$ and the covariance matrix $V_i^* = Z_i D^* Z_i + \sigma^{*2} I_{n_i}$ of the responses $\boldsymbol{y}_i$ will, asymptotically, be correctly estimated.

As an illustration of the previous theorem, we have simulated growth curves for 3 groups with an equal number of subjects in each group and such that the average profiles in each group are described by model (1) with parameter values as given in Table 1. The error terms were simulated from a normal distribution with mean zero and variance $\sigma^2 = 0.48$. The random effects $\boldsymbol{b}_i$ were sampled from one of the following distributions:

1. A normal distribution:

$$N\left[\mathbf{0}, \begin{pmatrix} 5.99 & -0.32 \\ -0.32 & 0.11 \end{pmatrix}\right].$$

2. A symmetric mixture of two normal distributions:

$$\frac{1}{2}N\left[\begin{pmatrix} -2 \\ 0.2 \end{pmatrix}, \begin{pmatrix} 1.99 & 0.08 \\ 0.08 & 0.07 \end{pmatrix}\right] + \frac{1}{2}N\left[\begin{pmatrix} 2 \\ -0.2 \end{pmatrix}, \begin{pmatrix} 1.99 & 0.08 \\ 0.08 & 0.07 \end{pmatrix}\right].$$

3. An asymmetric mixture of two normal distributions:

$$\frac{1}{10}N\left[\begin{pmatrix} -4.50 \\ 0.45 \end{pmatrix}, \begin{pmatrix} 3.7400 & -0.0950 \\ -0.0950 & 0.0875 \end{pmatrix}\right]$$
$$+\frac{9}{10}N\left[\begin{pmatrix} 0.50 \\ -0.05 \end{pmatrix}, \begin{pmatrix} 3.7400 & -0.0950 \\ -0.0950 & 0.0875 \end{pmatrix}\right].$$

4. Random intercepts and slopes, independently sampled from the lognormal distribution $LN(0,1)$ (see Johnson and Kotz (1970, pp. 112–136) and transformed such that the mean vector and the covariance matrix are the same as for the normal distribution in 1.

5. A discrete distribution with equal probability $\frac{1}{4}$ at the support points

$$\left\{\begin{pmatrix} -2.4479 \\ -0.1741 \end{pmatrix}, \begin{pmatrix} -2.4479 \\ 0.4356 \end{pmatrix}, \begin{pmatrix} 2.4479 \\ -0.4356 \end{pmatrix}, \begin{pmatrix} 2.4479 \\ 0.1741 \end{pmatrix}\right\}.$$

All distributions have been chosen such that $E(\boldsymbol{b}_i) = \mathbf{0}$ and such that $Var(\boldsymbol{b}_i)$ equals the random-effects covariance estimate given in Table 1. A graph of the bivariate density function for each of the above distributions is given in Fig. 1. For each random-effects distribution, datasets were simulated with total sample size N equal to 30, 60, 120, 240, 480, 960 and 1920, respectively. All these datasets have been analysed with a linear mixed-effects model with the same fixed and random-effect covariates as were used to analyse Goldstein's data. Normality was assumed for the random effects. The relative distances $\|\widehat{\boldsymbol{\xi}}_N - \boldsymbol{\xi}\|/\|\boldsymbol{\xi}\|$ of the fixed effects, the parameters in D and the residual variance separately, are shown in Fig. 2, as functions of the total sample size. These graphs clearly illustrate the consistency of the MLEs but also suggest that the rate of convergence heavily depends on the correct random-effects distribution (especially for the estimation of D).

Before discussing the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_N$, we first introduce some additional notation. Let $q_{ik}(\boldsymbol{\theta})$ and $q_{ikl}(\boldsymbol{\theta})$ denote the first and second order partial

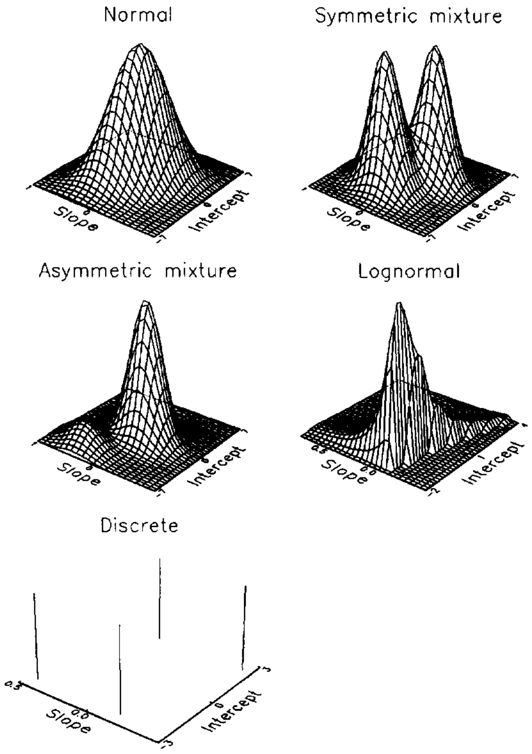

Fig. 1. The bivariate density function for each of the five distributions from which random effects were sampled in the simulation study.

derivatives of $\ln f_i(y_i|\boldsymbol{\theta})$ with respect to θ_k and with respect to θ_k and θ_l respectively.

We then define matrices

$$A_N(\boldsymbol{\theta}) = \left[-\frac{1}{N}\sum_{i=1}^{N} q_{ikl}(\boldsymbol{\theta})\right]_{k,l}, \qquad B_N(\boldsymbol{\theta}) = \left[\frac{1}{N}\sum_{i=1}^{N} q_{ik}(\boldsymbol{\theta})q_{il}(\boldsymbol{\theta})\right]_{k,l}$$

as well as their expectations $\bar{A}_N(\boldsymbol{\theta})$ and $\bar{B}_N(\boldsymbol{\theta})$. Expressions for $q_{ik}(\theta)$ and for $q_{ikl}(\theta)$ can be obtained from extending the formulas by Lindstrom and Bates (1988, Section 3.2) to the case where D is assumed to be symmetric, and are given by Verbeke and Lesaffre (1994). Further, let $\bar{A}_N^*$ and $\bar{B}_N^*$ be equal to $\bar{A}_N$ and $\bar{B}_N$, but evaluated at the correct parameter value θ^*. We then have the following theorem.

Theorem 2. *Under general regularity conditions, we have that* $\widehat{\boldsymbol{\theta}}_N$ *is asymptotically normally distributed with mean* $\boldsymbol{\theta}^*$ *and with asymptotic covariance matrix* $\bar{A}_N^{*-1}\bar{B}_N^*\bar{A}_N^{*-1}/N$, *for* $N \to \infty$.

Note that, when the assumed model is correct, we have that $\bar{A}_N^* = \bar{B}_N^*$. Theorem 2 then reduces to the classical result of asymptotic normality with the inverse Fisher

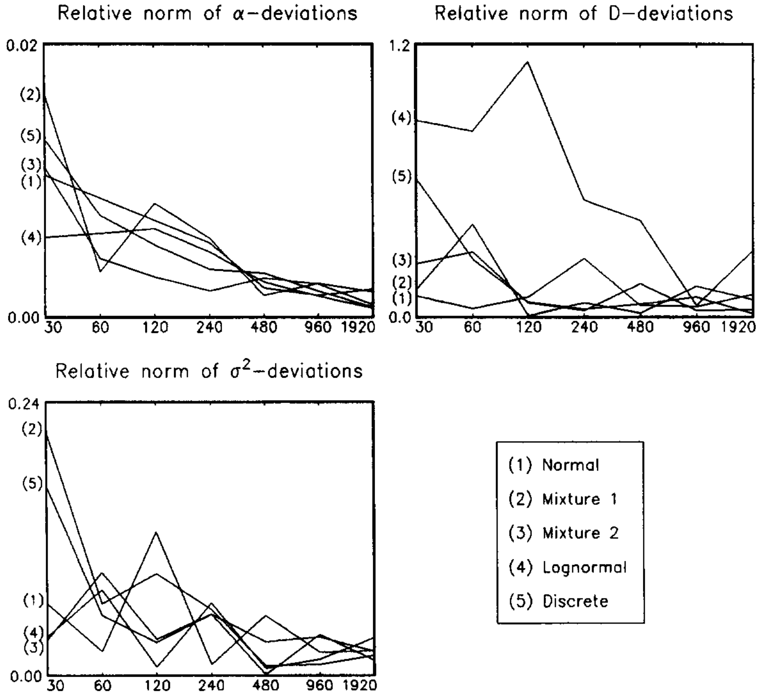

Fig. 2. Relative distances between parameters $\boldsymbol{\alpha}$, D and σ^2, and their maximum likelihood estimates $\widehat{\boldsymbol{\alpha}}_N$, $\widehat{D}_N$ and $\widehat{\sigma}_N^2$, versus the total sample size in simulated data with five different random-effects distributions.

information matrix as asymptotic covariance matrix for the vector of MLEs (see for example Fahrmeir and Kaufmann (1985) for the case of generalized linear models). Further, it can easily be shown that the second derivatives of $\ln f_i(\mathbf{y}_i|\boldsymbol{\theta}) = \ln f_i(\mathbf{y}_i|\boldsymbol{\alpha}, \sigma^2, D)$ with respect to $\boldsymbol{\alpha}$ and σ^2, and with respect to $\boldsymbol{\alpha}$ and the elements of D, have mean zero. Hence, under normality of the random effects, we have that the fixed-effects estimator $\widehat{\boldsymbol{\alpha}}_N$ is asymptotically independent of the variance components estimator $(\widehat{\sigma}_N^2, \widehat{D}_N)$.

4. Correction of standard errors

Since calculation of $\bar{A}_N^*$ and $\bar{B}_N^*$ requires knowledge of both the correct random-effects distribution and of $\boldsymbol{\theta}^*$, Theorem 2 cannot be used to test hypotheses about $\boldsymbol{\theta}^*$. Fortunately, the next theorem shows that a valid test can be obtained from replacing these matrices by their estimated versions $\widehat{A}_N$ and $\widehat{B}_N$ which are equal to A_N and B_N, evaluated at the MLE $\widehat{\boldsymbol{\theta}}$.

Theorem 3. *The results in Theorem 2 remain valid when $\bar{A}_N^*$ and $\bar{B}_N^*$ are replaced by $\widehat{A}_N$ and $\widehat{B}_N$ respectively.*

The asymptotic covariance matrix of estimators obtained with the maximum likelihood approach is usually estimated by the inverse of the matrix of second derivatives of the log-likelihood function ($=\widehat{A}_N^{-1}/N$), evaluated at the MLEs. The above theorem now tells us how this matrix should be corrected in cases where the random effects are possibly not normally distributed.

Let us now partition $\bar{A}_N^*$ and $\widehat{A}_N$ as

$$\bar{A}_N^* = \begin{pmatrix} \bar{A}_{N,11}^* & \bar{A}_{N,12}^* \\ \bar{A}_{N,21}^* & \bar{A}_{N,22}^* \end{pmatrix}, \qquad \widehat{A}_N = \begin{pmatrix} \widehat{A}_{N,11} & \widehat{A}_{N,12} \\ \widehat{A}_{N,21} & \widehat{A}_{N,22} \end{pmatrix},$$

where $\bar{A}_{N,11}^*$ ($\widehat{A}_{N,11}$) and $\bar{A}_{N,22}^*$ ($\widehat{A}_{N,22}$) are those submatrices which correspond to the fixed effects $\boldsymbol{\alpha}$ and the variance components σ^2 and D respectively. We then have that, under normality for the random effects, the asymptotic covariance matrix of $\widehat{\boldsymbol{\alpha}}_N$ can be estimated by the corresponding submatrix of $\widehat{A}_N^{-1}/N$, which is given by

$$\frac{1}{N}(\widehat{A}_{N,11}^{-1} + \widehat{A}_{N,11}^{-1}\widehat{A}_{N,12}(\widehat{A}_{N,22} - \widehat{A}_{N,21}\widehat{A}_{N,11}^{-1}\widehat{A}_{N,12})^{-1}\widehat{A}_{N,21}\widehat{A}_{N,11}^{-1}). \tag{2}$$

It follows from the asymptotic orthogonality of the estimators for the fixed effects and those for the variance components that $\bar{A}_{N,12}^* = 0$, and therefore that (2) can be approximated by

$$\frac{1}{N}\widehat{A}_{N,11}^{-1} = \left(\sum_{i=1}^{N} X_i'\widehat{V}_i^{-1}X_i\right)^{-1}. \tag{3}$$

Note that the above expression is frequently used as approximate covariance matrix for $\widehat{\boldsymbol{\alpha}}_N$ (see for example procedure MIXED in SAS (1992)), since it naturally arises from the fact that

$$\widehat{\boldsymbol{\alpha}}_N = \left(\sum_{i=1}^{N} X_i'\widehat{V}_i^{-1}X_i\right)^{-1} \sum_{i=1}^{N} X_i'\widehat{V}_i^{-1}\boldsymbol{y}_i.$$

Hence, we have that the asymptotic covariance matrix suggested by Theorem 3 adds extra variability to (3), due to the estimation of the variance components, but also corrects for possible misspecification of the random-effects distribution.

In order to investigate the effect of this last correction, we can compare the covariance matrix for the estimators obtained in Theorem 3 (= corrected) with the one obtained from the classical maximum likelihood theory (= uncorrected). The variance of any linear combination $\boldsymbol{\lambda}'\widehat{\boldsymbol{\psi}}_N$ is estimated by $\boldsymbol{\lambda}'\widehat{A}_N^{-1}\widehat{B}_N\widehat{A}_N^{-1}\boldsymbol{\lambda}/N$ and by $\boldsymbol{\lambda}'\widehat{A}_N^{-1}\boldsymbol{\lambda}/N$, respectively. It now follows from Seber (1984, p. 526) that, for all $\boldsymbol{\lambda}$, the ratio of both estimates satisfies

$$\lambda_{\min} \le \frac{\boldsymbol{\lambda}'\widehat{A}_N^{-1}\widehat{B}_N\widehat{A}_N^{-1}\boldsymbol{\lambda}}{\boldsymbol{\lambda}'\widehat{A}_N^{-1}\boldsymbol{\lambda}} \le \lambda_{\max}, \tag{4}$$

where $\lambda_{\min}$ is the smallest and $\lambda_{\max}$ the largest eigenvalue of $\widehat{B}_N\widehat{A}_N^{-1}$, and that $\lambda_{\min}$ ($\lambda_{\max}$) is attained for $\boldsymbol{\lambda}$ equal to an eigenvector for $\lambda_{\min}$ ($\lambda_{\max}$). So, $\lambda_{\min} = \lambda_{\max} = 1$

means that both inferences would yield the same results. Therefore $\lambda_{\min} \approx \lambda_{\max} \approx 1$ may indicate that the random effects are approximately normally distributed.

5. Finite sample results

In order to compare the performance of the uncorrected and corrected standard errors in finite samples, we have set up an extensive simulation study based on the same five models as were used for the simulations described in Section 3.

From each of these models, K samples (K will be specified below) were drawn, with group sizes 40, 80 and 120. Each of the K datasets is analysed with model (1) assuming normality for the random effects. Parameter estimates and uncorrected as well as corrected standard errors are calculated. Our model contains 6 fixed effects and 4 variance components. Because the results for the fixed effects are very similar for the 3 groups, we will only report the results for the average intercept α_1 and average slope α_4 for group A, and the results for the variance components.

The performances of the uncorrected standard errors and corrected standard errors were compared by evaluation of coverage probabilities of confidence intervals constructed with these standard errors. The proportion $\widehat{\pi}$ of 95% confidence intervals containing the correct parameter was calculated, together with the 95% confidence interval for the population proportion π. The results of this analysis are graphically shown on the left hand sides of the Figs. 3–7, corresponding to the 5 random-effects distributions used in the simulations. Each panel corresponds to a specific parameter in the model and contains the 95% confidence intervals for π, based on the uncorrected (left hand side interval) and the corrected (right hand side interval) standard errors, for the three different group sizes (40, 80, 120). In the absence of bias, and when the standard errors are correct, these confidence intervals of π should contain the value 0.95. The horizontal line in the graphs represents this reference value. In order to get confidence intervals for π with length smaller than 0.02 when $\widehat{\pi} = 0.95$, we have chosen $K = 2000$.

Estimated standard errors were compared with the sample standard deviations from the 2000 simulated trials as well. The results are shown in the right hand sides of Figs. 3–7. Each of the panels contains a boxplot of the uncorrected (left hand side plot) and the corrected (right hand side plot) standard errors, for the three different group sizes (40, 80, 120). These boxplots show the median and the 2.5%, 25%, 75% and 97.5% percentiles of the distributions. Good standard errors are those with small variability around the sample deviation. The horizontal lines in the graphs represent these reference values.

In general, we can conclude that, for the fixed effects, the corrected and uncorrected standard errors are very similar. For the random components on the other hand, and more specifically for the elements of the random-effects covariance matrix D, this is only true under the correct model (normal random effects, Fig. 3). When the random effects are not normally distributed, we found that the corrected standard errors are clearly superior to the uncorrected ones, but occasionally still not performing very well.

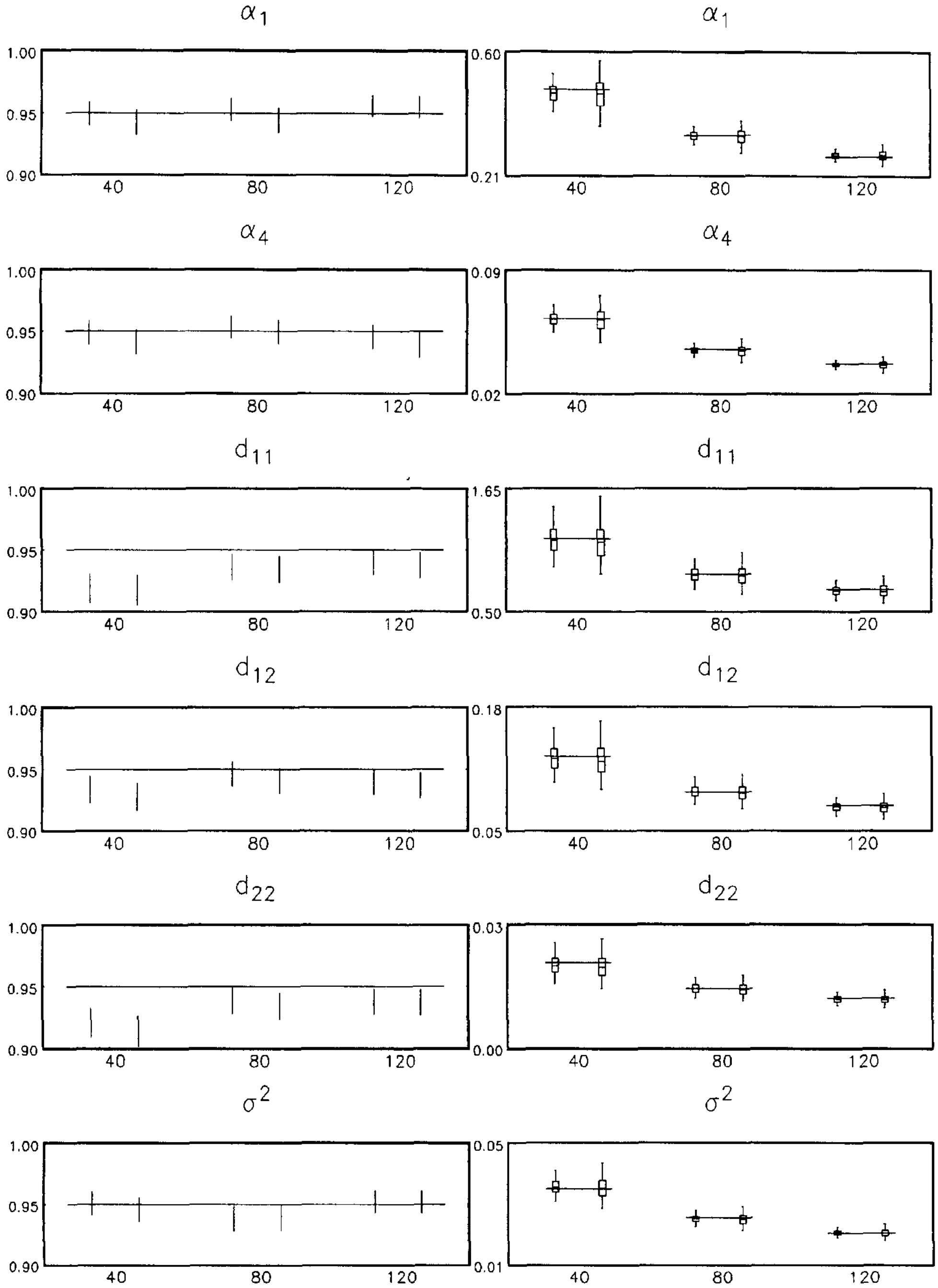

Fig. 3. Summary of the simulation results for the case where the random effects are normally distributed. The left hand figures show 95% confidence intervals for the probability that a 95% confidence interval (uncorrected and corrected) contains the true parameter value. The horizontal line indicates the pursued confidence level 95%. The right hand figures contain boxplots of the estimated (uncorrected and corrected) standard errors. The horizontal lines correspond to the sample standard deviations. All of this is done for group sizes 40, 80 and 120.

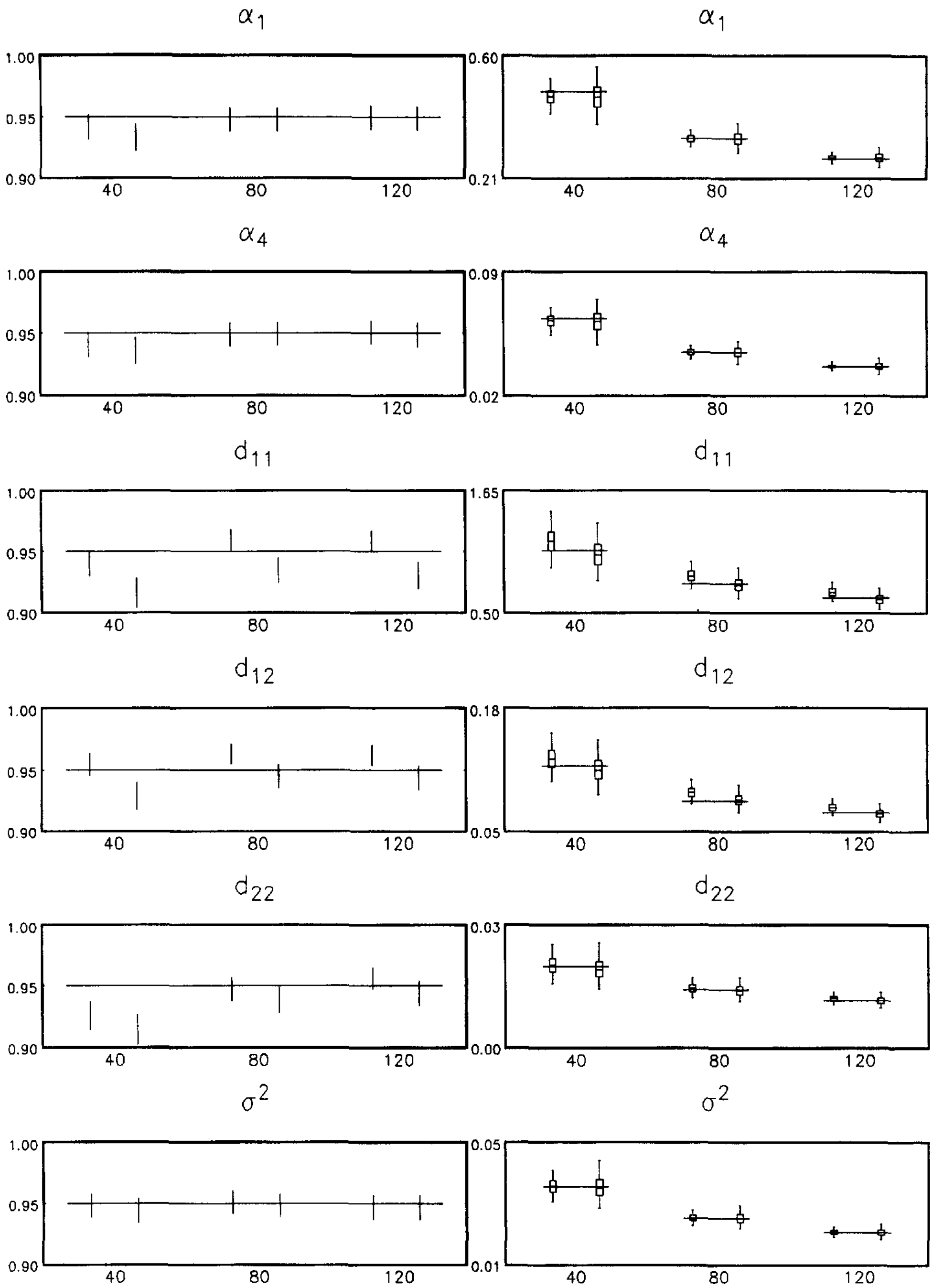

Fig. 4. Summary of the simulation results for the case where the random effects are sampled from a symmetric mixture of two normal distributions. The left hand figures show 95% confidence intervals for the probability that a 95% confidence interval (uncorrected and corrected) contains the true parameter value. The horizontal line indicates the pursued confidence level 95%. The right hand figures contain boxplots of the estimated (uncorrected and corrected) standard errors. The horizontal lines correspond to the sample standard deviations. All of this is done for group sizes 40, 80 and 120.

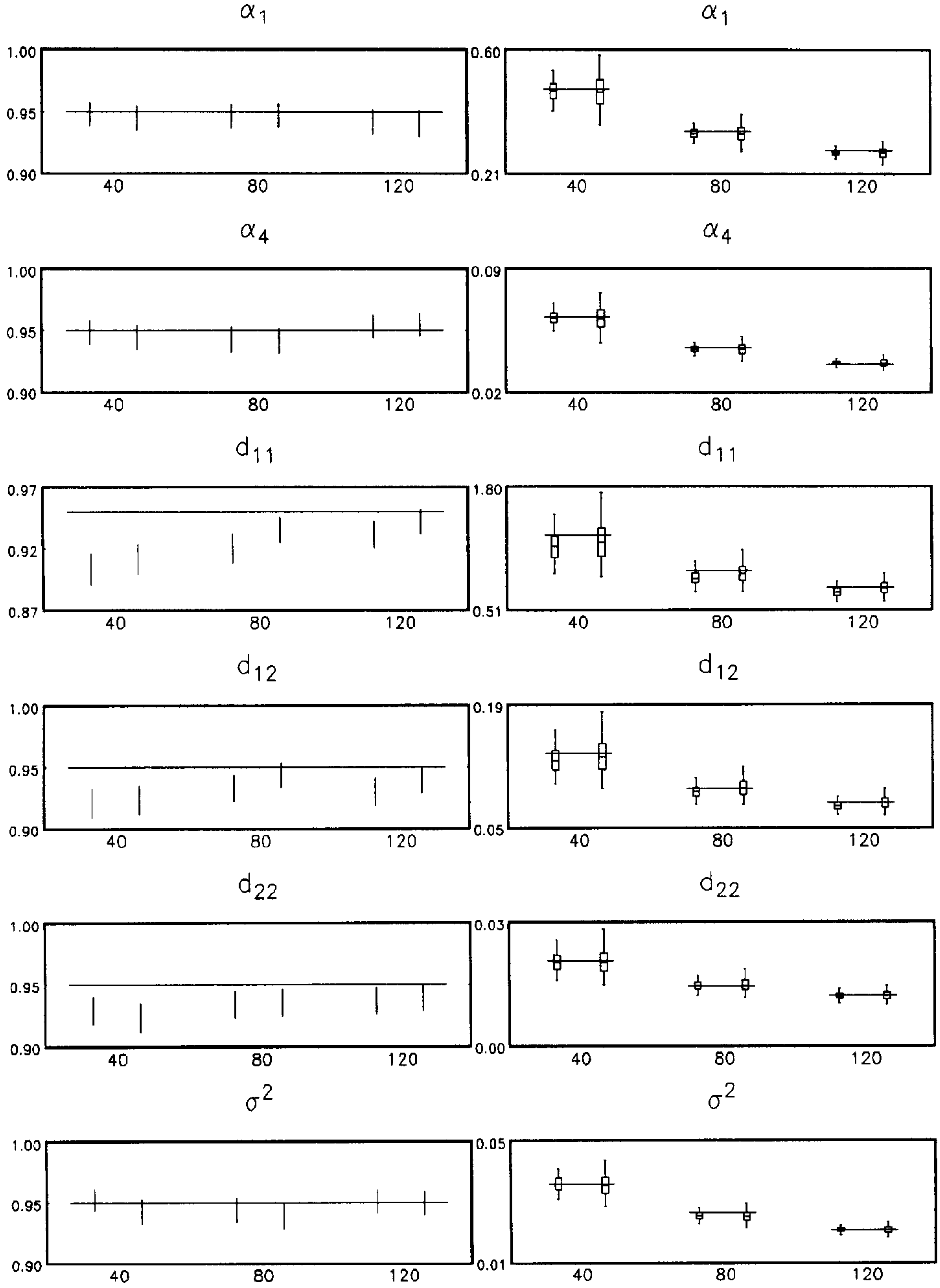

Fig. 5. Summary of the simulation results for the case where the random effects are sampled from an asymmetric mixture of two normal distributions. The left hand figures show 95% confidence intervals for the probability that a 95% confidence interval (uncorrected and corrected) contains the true parameter value. The horizontal line indicates the pursued confidence level 95%. The right hand figures contain boxplots of the estimated (uncorrected and corrected) standard errors. The horizontal lines correspond to the sample standard deviations. All of this is done for group sizes 40, 80 and 120.

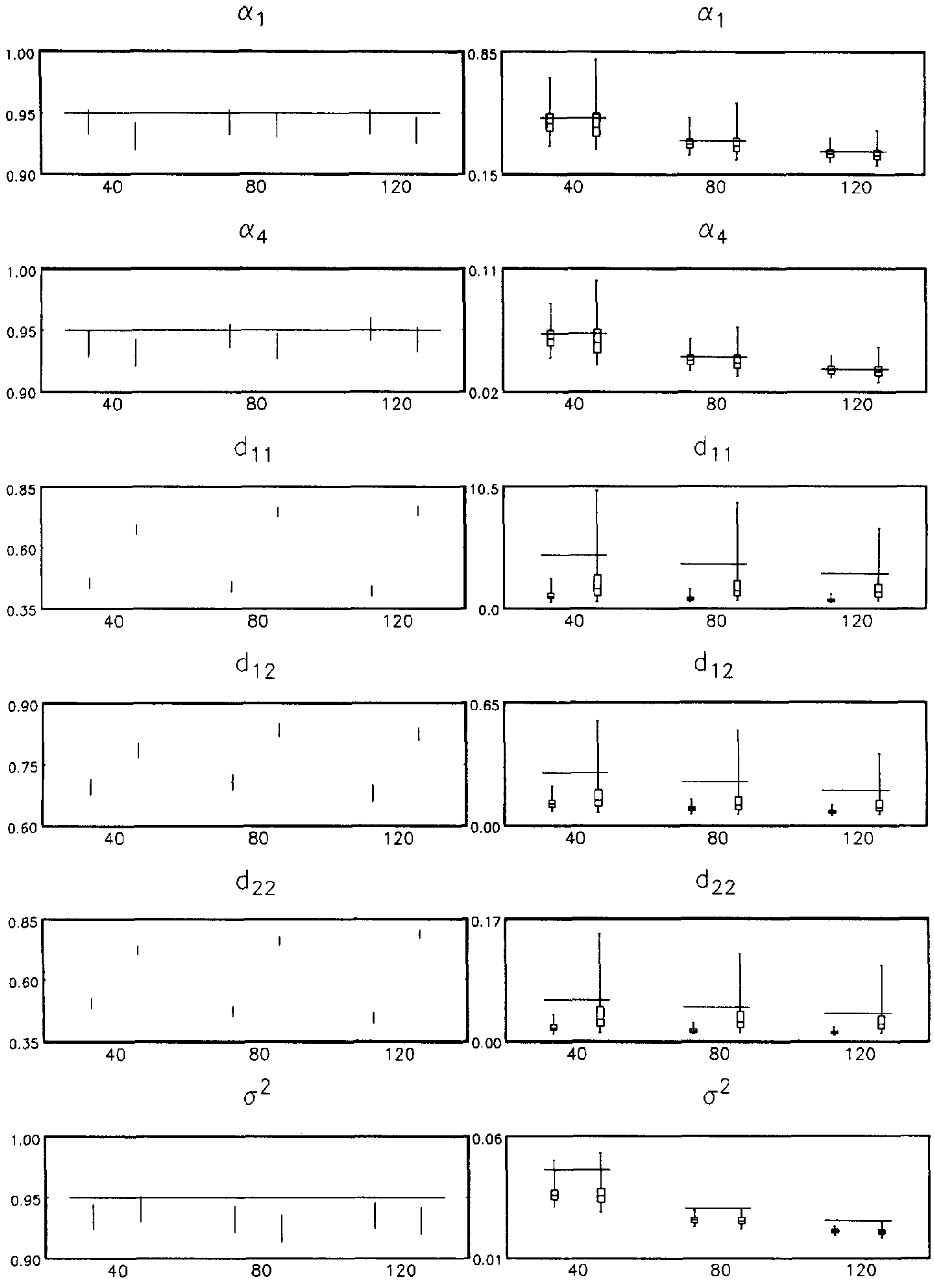

Fig. 6. Summary of the simulation results for the case where the random effects are log-normally distributed. The left hand figures show 95% confidence intervals for the probability that a 95% confidence interval (uncorrected and corrected) contains the true parameter value. The horizontal line indicates the pursued confidence level 95%. The right hand figures contain boxplots of the estimated (uncorrected and corrected) standard errors. The horizontal lines correspond to the sample standard deviations. All of this is done for group sizes 40, 80 and 120.

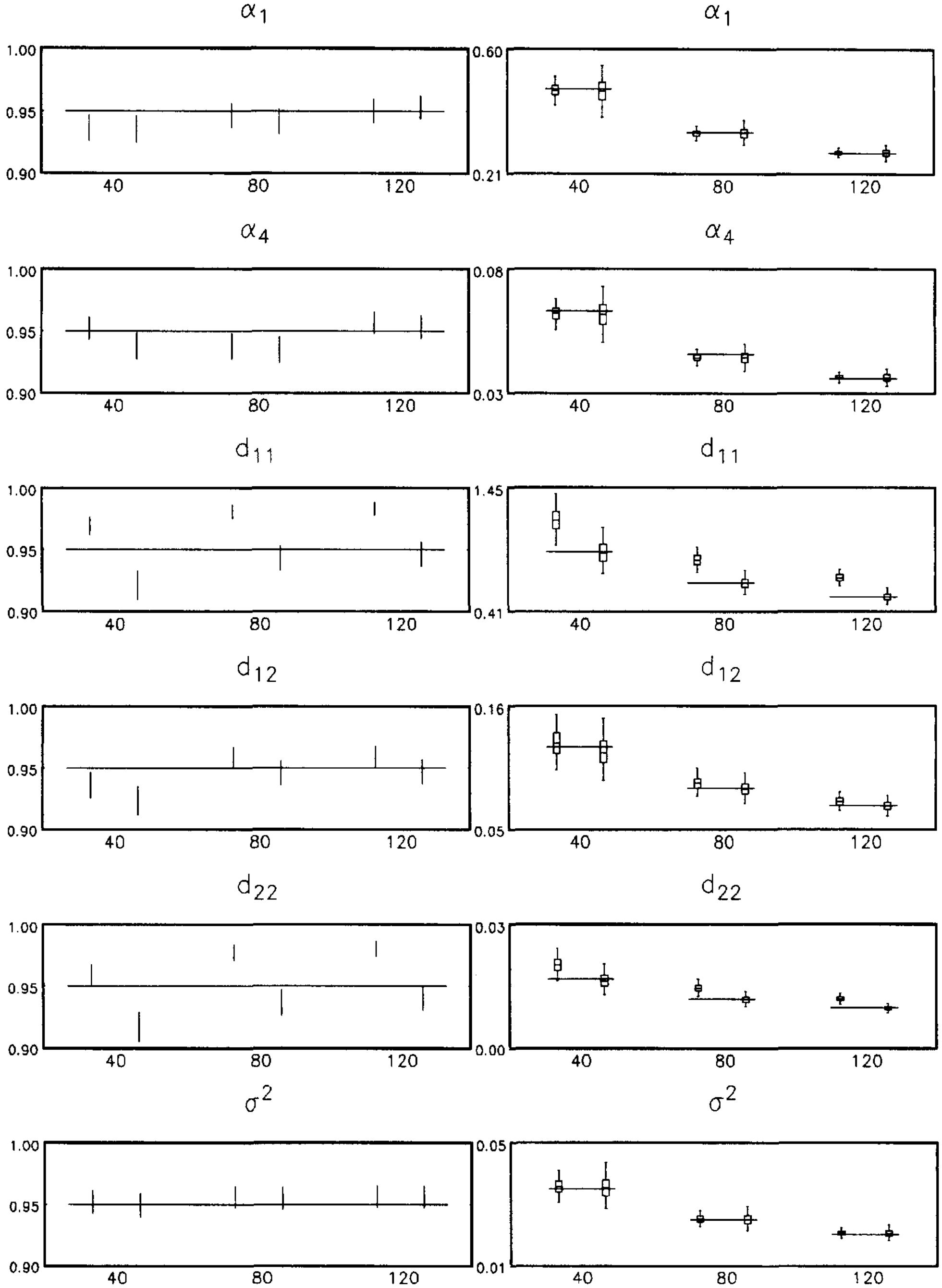

Fig. 7. Summary of the simulation results for the case where the random effects are sampled from a discrete distribution. The left hand figures show 95% confidence intervals for the probability that a 95% confidence interval (uncorrected and corrected) contains the true parameter value. The horizontal line indicates the pursued hand figures contain boxplots of the estimated (uncorrected and corrected) standard errors. The horizontal lines correspond to the sample standard deviations. All of this is done for group sizes 40, 80 and 120.

In some cases (see for example Figs. 5 and 6), the correction enlarges the standard errors to get larger intervals with confidence levels closer to the pursued level. For example, in 99.95% of the datasets with group size 120, and with lognormally distributed random effects, the uncorrected confidence interval for d_{11} was smaller than the corrected one, and therefore contained the correct parameter value less often. Further, $\lambda_{\max}$ was larger than 10 in 1217 out of the 2000 datasets, while $\lambda_{\min}$ was only 72 times smaller than 0.2.

In other cases (Fig. 7), the correction results in smaller standard errors which protect against confidence intervals which are unnecessary large and therefore too conservative. For example, for all 2000 simulated datasets with group size 120 and with discrete random effects, the uncorrected confidence interval for d_{11} was larger than the confidence interval based on the corrected standard error, leading to confidence intervals with confidence level larger than 95%. We also found that in this case, $\lambda_{\max}$ was always smaller than 1.6 while $\lambda_{\min}$ was smaller than 0.2 in 70.8% of the datasets.

Although the corrected standard errors are good estimates for the variability of the parameter estimators, they may still yield incorrect confidence intervals for small samples, due to biased estimation of the parameters. This is especially the case for the random components in the model (see for example Fig. 4).

Note that $\widehat{A}_N^{-1}\widehat{B}_N\widehat{A}_N^{-1}/N$ is of the same form as the so-called 'information sandwich' estimator for the asymptotic covariance matrix of fixed effects, estimated with quasi-likelihood methods (see e.g. Liang and Zeger, 1986; Firth, 1993). However, our asymptotic result relates to both the fixed effects and the parameters in the 'working correlation' model. Further, our model is incorrectly specified only through the random-effects distribution; the covariance structure is assumed to be correct. In the context of correlated binary data, Sharples and Breslow (1992) have shown that the sandwich estimator $\widehat{A}_N^{-1}\widehat{B}_N\widehat{A}_N^{-1}/N$ for the covariance matrix of fixed effects is almost as efficient as the uncorrected, model based estimator $\widehat{A}_N^{-1}/N$, when the assumed form of the covariance matrix is correct, even under the correct model. This is in agreement with our simulation results, reported in Fig. 3. Finally, research on the extension of our results to the class of generalized linear models seems promising, keeping in mind the empirical results of Butler and Louis (1992) for the case of binary data.

References

BMDP Statistical Softwar, Inc., 1440 Sepulveda Boulevard Suite 316, Los Angeles, CA 90025, *BMDP Statistical Software Manual*, Vol. 2 (1990).

Butler, S.M. and T.A. Louis, Random effects models with non-parametric priors, *Statist. Med.*, **11** (1992) 1981–2000.

Fahrmeir, L. and H. Kaufmann, Consistency and asymptotic normality of the maximum likelihood estimator in generalized models, *Ann. Statist.*, **13** (1985) 342–368.

Firth, D., Recent developments in quasi-likelihood methods, in: *Proc. ISI 49th Session*, Firenze (1993) 341–358.

Goldstein H., *The design and analysis of longitudinal studies. Their role in the measurement of change* (Academic Press, New York, 1979).

Golub G.H. and C.F. Van Loan, *Matrix computations*, 2nd edn. (Johns Hopkins University press, Baltimore, MD, 1989).

Johnson, N.L. and S. Kotz, *Continuous univariate distributions* – 1 (Houghton Mifflin, Boston, MA, 1970).

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

Lindstrom, M.J. and D.M. Bates, Newton–Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *J. Amer. Statist. Assoc.*, **83** (1988) 1014–1022.

SAS Institute Inc., Cary, NC, *SAS Technical Report P-229*, *SAS/STAT Software: Changes and Enhancements*, *Release* 6.07, 1992.

Seber, G.A.F., *Multivariate observations* (Wiley, New York, 1984).

Sharples, K. and N.E. Breslow, Regression analysis of correlated binary data: some small sample results for the estimating equation approach, *J. Statist. Comput. Simulation*, **42** (1992) 1–20.

Verbeke, G. and E. Lesaffre, Large sample properties of the maximum likelihood estimators in linear mixed models with misspecified random-effects distributions, Technical Report, Report #1996.1 Biostatistical Centre for Clinical Trials, Catholic University of Leuven, Belgium, 1994.

Verbeke, G. and E. Lesaffre, A linear mixed-effects model with heterogeneity in the random-effects population, *J. Amer. Statist. Assoc.*, **91** (1996) 217–221.