

A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models

Peer-reviewed author version

ALONSO ABAD, Ariel; LITIERE, Saskia & MOLENBERGHS, Geert (2008) A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models. In: COMPUTATIONAL STATISTICS & DATA ANALYSIS, 52(9). p. 4474-4486.

DOI: 10.1016/j.csda.2008.02.033

Handle: <http://hdl.handle.net/1942/8057>

A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models

A. Alonso ^a, S. Litière ^{a,*}, G. Molenberghs ^a

^a*Center for Statistics, Hasselt University, Agoralaan 1, B-3590 Diepenbeek,
Belgium*

Abstract

Estimation in generalized linear mixed models for non-Gaussian longitudinal data is often based on maximum likelihood theory, which assumes that the underlying probability model is correctly specified. It is known that the results obtained from these models are not always robust against misspecification of the random-effects structure. Therefore, diagnostic tools for the detection of this misspecification are of the utmost importance. Three diagnostic tests, based on the eigenvalues of the variance-covariance matrices for the fixed-effects parameters estimates, are proposed in the present work. The power and type I error rate of these tests are studied via simulations. A very acceptable performance was observed in many cases, especially for those misspecifications that can have a big impact on the maximum likelihood estimators.

Key words: Generalized linear mixed models; Misspecification; Power; Type I error;

1 Introduction

The generalized linear mixed model (GLMM) is a frequently used tool in the analysis of discrete repeated measurements (Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993; Diggle et al., 2002; Molenberghs and Verbeke, 2005). It is a natural extension of the generalized linear model, taking into account multiple sources of variation. Indeed, let y_{ij} denote the j th response of subject i , $i = 1, \dots, n$ and $j = 1, \dots, n_i$. Conditional on a vector of individual random effects \mathbf{b}_i , the outcome variables are assumed to be independent with density functions belonging to the exponential family

$$f(y_{ij}|\mathbf{b}_i, \theta_{ij}, \varphi) = \exp[\varphi^{-1}\{y_{ij}\theta_{ij} - \psi(\theta_{ij})\} + c(y_{ij}, \varphi)], \quad (1)$$

where φ is a scale parameter, $c(\cdot)$ is a function only depending on y_{ij} and φ , and $\psi(\cdot)$ is a function satisfying $E(y_{ij}|\mathbf{b}_i) = \psi'(\theta_{ij})$ and $\text{Var}(y_{ij}|\mathbf{b}_i) = \varphi\psi''(\theta_{ij})$. Further, $\mu_{ij} = E(y_{ij}|\mathbf{b}_i) = v(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i)$, where $v(\cdot)$ denotes a known link function, \mathbf{x}_{ij} and \mathbf{z}_{ij} are vectors of covariates, and $\boldsymbol{\beta}$ is a vector of unknown fixed regression coefficients. The subject-specific effects \mathbf{b}_i are often assumed to be normally distributed with mean zero and variance-covariance matrix D . Fitting the model requires maximization of the marginal likelihood, which is obtained by integrating out the random effects. Let the contribution to the likelihood of subject i be given by:

$$f(\mathbf{y}_i|\boldsymbol{\beta}, D, \varphi) = \int \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \theta_{ij}, \varphi) f(\mathbf{b}_i|D) d\mathbf{b}_i, \quad (2)$$

* Corresponding author. Tel.: +32 11 268282; fax: +32 11 268205.
Email address: saskia.litiere@uhasselt.be (S. Litière).
URL: <http://www.uhasselt.be/censtat> (S. Litière).

then we can write the marginal likelihood as

$$L(\boldsymbol{\beta}, D, \varphi) = \prod_{i=1}^n \int \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{b}_i, \theta_{ij}, \varphi) f(\mathbf{b}_i | D) d\mathbf{b}_i. \quad (3)$$

The commonly used choice of the normal distribution for these random effects generally leads to intractable likelihood functions, with the exception of linear mixed model (LMM) for continuous data. In response, several numerical approximations to the likelihood have been implemented in the available software packages. For example, Gaussian quadrature, as implemented in the SAS procedure NLMIXED, approximates the integral using Gaussian-Hermite polynomials, thereby employing specific properties of the normal distribution.

Obviously, estimation and inferences based on this model, depend on the assumption that the model, and therefore the random-effects structure, is correctly specified. Since random effects are unobserved, violations of the assumptions concerning this part of the model may be difficult to identify. Recent research has focused on the impact of misspecifications of the random effects structure on the maximum likelihood estimators and other inferential procedures. [Neuhaus et al. \(1992\)](#) showed that the maximum likelihood estimators of a random-intercept logistic model with misspecified random-effect distribution are inconsistent, but that the magnitude of the bias is typically small. Simulations by [Chen et al. \(2002\)](#) also indicate that the estimation of the regression coefficients may be subject to negligible bias when the distribution of the random effects is wrongly assumed to be normal. According to [Agresti et al. \(2004\)](#), the choice of the random-effects distribution seems to have, in most situations, little effect on the maximum likelihood estimators. However, when there is a severe polarization of subjects, for example, by omitting an influential binary covariate, this can affect the predictive qualities of characteristics

involving the random effects as well as the fixed effects. Similarly, [Heagerty and Kurland \(2001\)](#) found substantial bias while using a random-intercept model when the random-effect distribution depends on measured covariates.

We should underscore that all these simulation studies were performed using a limited number of distributions and, in all of them, only small variances for the random-effects were considered. [Litière et al. \(2008\)](#) found, using simulations with a random-intercept logistic model and a wide range of distributions for the random effect, that the estimate of the variance component is always subject to considerable bias when the random-effect distribution is misspecified. Although variance components are generally treated as nuisance parameters, this bias can have an important impact in studies where they are of primary interest. This is the case, for instance, in fields like surrogate marker validation, the assessment of the reliability of rating scales, or studies to evaluate the criterion and predictive validity of psychiatric rating scales.

Furthermore, the bias induced in the estimates of the mean structure parameters appears to depend on the magnitude of the variance component, whereby a large bias is associated with large random-effects variance. Clearly, in any practical situation, the bias induced in the estimator of the variance component under misspecification will make it hard to distinguish between the two scenarios, that is, small or larger variance components. Therefore, it can be difficult to determine how severe the impact on the mean parameters can be. Additionally, [Litière et al. \(2007\)](#) established that the type I error rate and the power related to the tests for the mean structure parameters, can also be severely affected, even in the presence of small variance component. These authors also showed with a limited simulation study that the situation worsens when more complicated random-effects structures are used.

In these circumstances, the development of diagnostic tools that can help to detect these kinds of misspecifications is of great importance. [White \(1982\)](#) proposed a general test for model misspecification, the so-called Information-Matrix Test (IMT). Unfortunately, this test requires third-order partial derivatives of the likelihood function. Even though the calculation of higher order derivatives might not be an issue in cases where the likelihood is available in a closed form, as for example, in the LMM case, it can become an important problem when such a closed form does not exist, as for example, in the GLMM case. More recently, [Waagepetersen \(2006\)](#) proposed a simulation-based test by generating random effects while conditioning on the observations. The intuition behind this test is that, if the joint model for the observations and the random effects is correctly specified, then the marginal distribution of the simulated random effects should coincide with the assumed distribution. Although simulations with Poisson responses showed a reasonable power, this test required very large cluster and sample sizes to produce similar results with binary outcomes. [Tchetgen and Coull \(2006\)](#) proposed a diagnostic test to verify the validity of the choice of the random-effects distribution by comparing marginal and conditional maximum likelihood estimators of a subset of fixed effects in the model. They argue that the conditional estimators are robust to the choice of the random-effects distribution, whereas the estimators from the marginal model will be affected if this distribution is misspecified. Therefore, they suggest a test statistic based on the difference between these estimates, focusing on the covariates which vary within each cluster. Clearly, the test is restricted to those applications which involve at least one within-cluster covariate. This would make it inapplicable, for instance, to study the appropriateness of the normal distribution to describe the heterogeneity of the latent trait involved in the Rasch model and other item response mod-

els (Agresti, 2002). Another limitation of this test is that it cannot be applied when auto-regressive random effects are present. However, as we will illustrate in the following sections, misspecification in this scenario can have serious consequences for the inferences and the estimation procedures.

In the present work, we extend this toolbox with a set of diagnostic tools based on the eigenvalues of the variance-covariance matrix of the parameters estimators.

In Section 2 the motivating case study is introduced. In Section 3, we derive the new tests. Next, in Section 4, the performance of these proposals is evaluated by means of simulations. Finally, the appropriateness of the model chosen to analyze the case study is assessed, using the new tests, in Section 5.

2 A Case Study in Mental Health

The data were obtained from a clinical trial investigating the effect of risperidone compared to an active control for the treatment of chronic schizophrenia (Alonso et al., 2004). In total, 128 patients were included in the trial, from which 64 were assigned to the treatment group ($z_i = 1$) and the rest to the control group ($z_i = 0$). Several measures can be used to assess a patient's global condition. The *Clinical Global Impression* (CGI) is generally accepted as a clinical measure of change. It is a 7-grade scale used to characterize a subject's mental condition. Clinicians are often interested in a transformation of this scale into a binary response variable Y which equals 1 for patients classified as normal to mildly ill, and 0 for patients classified as moderately to severely ill. In the study treatment was administered for 8 weeks and the

Table 1

Case study. Parameter estimates (standard errors) and p-values for the effects in the logistic random-intercept model (4).

Effect	Parameter	Estimate (s.e.)	p-value
Fixed effects			
Intercept	β_0	-7.37 (1.18)	< 0.0001
Treatment effect	β_1	2.14 (1.08)	0.049
Time effect	β_2	0.65 (0.10)	< 0.0001
Variance structure			
Random intercept variance	σ_b^2	21.01 (6.81)	

outcome was measured at 6 fixed time points: 0, 1, 2, 4, 6, and 8 weeks.

A previous model building endeavor led to the following model for these data ([Litière et al., 2007](#)):

$$\text{logit}\{P(y_{ij} = 1|b_{0i})\} = \beta_0 + b_{0i} + \beta_1 z_i + \beta_2 t_j, \quad (4)$$

where y_{ij} denotes the response for patient i at time point t_j , z_i denotes the treatment group to which subject i was allocated, and b_{0i} represents a random intercept, assumed to follow a mean-zero normal distribution with variance σ_b^2 . The corresponding maximum likelihood estimates are summarized in Table 1.

According to the results obtained by [Litière et al. \(2008\)](#), the relatively large estimate for the variance of the random effect reported in Table 1 could potentially imply a serious bias for the mean parameters, including the treatment effect, if the random-effect distribution is misspecified. In practice, such a large random-effect variance would not be rare, for instance, in clinical trials involving a placebo control, where little variability in the response is to be expected in the placebo group. For the data at hand, this large estimate could be explained by the high proportion of patients (75%) in the control group that have a response pattern consisting solely of zeros. Hence, two variance

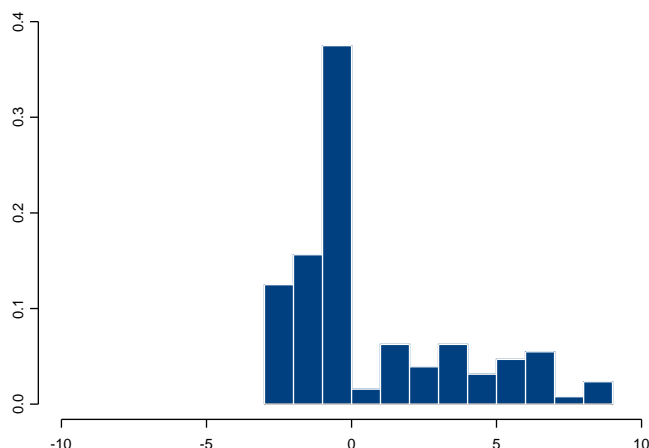


Fig. 1. *Case study. Histograms of the Empirical Bayes estimates obtained from fitting model (4), under the assumption of normally distributed random effects.*

components, one for each treatment group, might be necessary to capture this structure. However, allowing σ_b^2 to vary among the treatment groups did not decrease the variance components. Indeed, this analysis resulted in a random-effect variance of 20.00 (s.e. 7.93) for the treatment group and 22.61 (s.e. 10.69) for the control group. These high variances hint on a very strong intra-subject correlation within each treatment group.

Arguably, these circumstances could render the assumption of a normal distribution for the random effects questionable. The problem is aggravated by the unobservable nature of the random effects, complicating the evaluation of their distributional assumptions. Often, histograms of the empirical Bayes (EB) estimates of the random effects are used to detect departures from normality. For instance, in Figure 1, we have displayed the EB estimates obtained from fitting model (4) to the data of the case study. Observe that this histogram shows a severe departure from normality. Nevertheless, one should be careful in using these estimates to detect misspecification of the random-effects dis-

tribution, since it can be shown that in GLMM, the EB estimates no longer follow a normal distribution, even when the random-effects distribution is correctly specified as being of a normal type (Molenberghs and Verbeke, 2005). Another important issue occurring in this study was the high proportion of drop-out present in the last two measurement occasions, especially in week 8 for the control group (50%). A full analysis and discussion of this missing data problem is beyond the scope of the present work. Therefore, here we will assume that the missing data generating mechanism is missing at random (MAR) making our likelihood approach a valid option (Molenberghs and Verbeke, 2005; Kenward and Molenberghs, 2007). Note that MAR is oftentimes considered a plausible assumption in a clinical trial setting.

From the previous discussion it becomes clear that diagnostic tools which can detect the presence of this type of misspecification are of the utmost importance. In what follows, we will introduce three such tests, based on the eigenvalues of the variance-covariance matrix of the maximum likelihood estimators.

3 Methodology

Let us consider a random variable \mathbf{y} with density function h , and a parametric family of density functions $\mathfrak{F} = \{f(\mathbf{y}; \boldsymbol{\xi}) : \boldsymbol{\xi} \in \Upsilon\}$. In this manuscript, f denotes the marginal model (2) associated with the hierarchical model defined in (1), with random effects assumed to follow a normal distribution. Further, let $\boldsymbol{\xi}$ be a vector of order $p \times 1$, representing all parameters in (1), including the mean structure parameters $\boldsymbol{\beta}$ and the variance components in D .

In the absence of misspecification, i.e., if there exists a $\boldsymbol{\xi}_0 \in \Upsilon$ such that $h(\mathbf{y}) = f(\mathbf{y}, \boldsymbol{\xi}_0)$, it can be shown that the maximum likelihood estimator $\widehat{\boldsymbol{\xi}}_n$ is a consistent estimator for $\boldsymbol{\xi}_0$ and the inverse of Fisher's information matrix yields the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\xi}}_n$. However, this matrix does not provide valid results when the model is misspecified, and appropriate standard error estimates must be obtained by replacing the standard asymptotic covariance matrix by the so-called sandwich estimator (White, 1982). Let us further introduce the following additional notation

$$\begin{aligned}\mathbf{A}(\boldsymbol{\xi}) &= \mathbb{E} \left\{ \frac{\partial^2 \log f(\mathbf{y}_i, \boldsymbol{\xi})}{\partial \xi_k \partial \xi_\ell} \right\}, \\ \mathbf{B}(\boldsymbol{\xi}) &= \mathbb{E} \left\{ \frac{\partial \log f(\mathbf{y}_i, \boldsymbol{\xi})}{\partial \xi_k} \cdot \frac{\partial \log f(\mathbf{y}_i, \boldsymbol{\xi})}{\partial \xi_\ell} \right\}, \\ \mathbf{A}_n(\boldsymbol{\xi}) &= \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(\mathbf{y}_i, \boldsymbol{\xi})}{\partial \xi_k \partial \xi_\ell} \right\}, \\ \mathbf{B}_n(\boldsymbol{\xi}) &= \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(\mathbf{y}_i, \boldsymbol{\xi})}{\partial \xi_k} \cdot \frac{\partial \log f(\mathbf{y}_i, \boldsymbol{\xi})}{\partial \xi_\ell} \right\},\end{aligned}$$

where $k, \ell = 1, \dots, p$, and p refers to the number of parameters in the model. White (1982) proved that, under a correctly specified model, $\mathbf{B}(\boldsymbol{\xi}_0) + \mathbf{A}(\boldsymbol{\xi}_0) = \mathbf{0}$ and therefore, deviations from the model assumptions are expected to distort this equality. In what follows, we will use this property to construct tests to detect model misspecifications.

3.1 The Determinant Tests

Consider for each subject i the score statistic given by:

$$\mathbf{S}(\mathbf{y}_i, \boldsymbol{\xi}) = \left\{ \frac{\partial \log f(\mathbf{y}_i, \boldsymbol{\xi})}{\partial \xi_k} \right\}. \quad (5)$$

Further, assuming that $\mathbf{S}(\mathbf{y}_i, \boldsymbol{\xi}_0) \sim N_p[\mathbf{0}, -\mathbf{A}(\boldsymbol{\xi}_0)]$ and taking into account that $n\mathbf{B}_n(\boldsymbol{\xi}_0) = \sum_i \mathbf{S}(\mathbf{y}_i, \boldsymbol{\xi}_0)\mathbf{S}^T(\mathbf{y}_i, \boldsymbol{\xi}_0)$, the results established by [Anderson \(1963\)](#) and [Girschick \(1939\)](#) on the large sample distribution theory for the eigenvalues of a variance-covariance matrix show that, if $\gamma_1, \dots, \gamma_p$ represent the eigenvalues of $-\mathbf{A}(\boldsymbol{\xi}_0)$ and $\hat{\gamma}_{1n}, \dots, \hat{\gamma}_{pn}$ the eigenvalues of $\mathbf{B}_n(\boldsymbol{\xi}_0)$, then asymptotically

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}) \sim N_p(\mathbf{0}, 2\boldsymbol{\Gamma}^2), \quad (6)$$

where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^T$, $\hat{\boldsymbol{\gamma}}_n = (\hat{\gamma}_{1n}, \dots, \hat{\gamma}_{pn})^T$, and $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$. Applying the delta method, (6) can be transformed to $\sqrt{n}(\log \hat{\boldsymbol{\gamma}}_n - \log \boldsymbol{\gamma}) \sim N_p(\mathbf{0}, 2\mathbf{I})$, where \mathbf{I} is the $p \times p$ identity matrix. Using these results, we get:

Theorem 1 *Let us define $\delta_{d1}(n) = \log |\mathbf{B}_n(\boldsymbol{\xi}_0)[-\mathbf{A}_n^{-1}(\boldsymbol{\xi}_0)]|$ and $\delta_{d2}(n) = |\mathbf{B}_n(\boldsymbol{\xi}_0)| \cdot |-\mathbf{A}_n^{-1}(\boldsymbol{\xi}_0)|$. Then, under general regularity conditions, if the model is correctly specified*

$$1. \frac{n}{2p}[\delta_{d1}(n)]^2 \sim \chi_1^2, \quad \text{and} \quad (7)$$

$$2. \frac{n}{2p}[\delta_{d2}(n) - 1]^2 \sim \chi_1^2. \quad (8)$$

Given that $\mathbf{A}_n(\boldsymbol{\xi}_0)$ is a consistent estimator of $\mathbf{A}(\boldsymbol{\xi}_0)$, the asymptotic distribution of $\delta_{d1}(n)$ and $\delta_{d2}(n)$ will be the same as the asymptotic distribution of $\delta_{d1}^*(n) = \log |\mathbf{B}_n(\boldsymbol{\xi}_0)[-\mathbf{A}^{-1}(\boldsymbol{\xi}_0)]|$ and $\delta_{d2}^*(n) = |\mathbf{B}_n(\boldsymbol{\xi}_0)| \cdot |-\mathbf{A}^{-1}(\boldsymbol{\xi}_0)|$, respectively ([Lehmann, 1999](#)). Note now that $\delta_{d1}^*(n)$ can be written as $\delta_{d1}^*(n) = \sum_k (\log \hat{\gamma}_{kn} - \log \gamma_k)$. Since for each $k = 1, \dots, p$ the distribution of $\log \hat{\gamma}_{kn}$ is given by $(\log \hat{\gamma}_{kn} - \log \gamma_k) \sim N(0, 2/n)$, it can then easily be shown that $\delta_{d1}^*(n) \sim N(0, 2p/n)$, which is equivalent to (7). Furthermore, the second statistic equals $\delta_{d2}^*(n) = \prod_k \hat{\gamma}_{kn}/\gamma_k$. The variance of this statistic follows from the

distribution of $\widehat{\gamma}_{kn}$ in (6) after applying the delta method. This results in $\delta_{d2}^*(n) \sim N(1, 2p/n)$ and this implies that $\frac{n}{2p}[\delta_{d2}(n) - 1]^2 \sim \chi_1^2$.

In practical applications, $\boldsymbol{\xi}_0$ can be replaced by its consistent estimator under the null, given by $\widehat{\boldsymbol{\xi}}_n$. Further, note that, $\delta_{d1}(n)$ and $\delta_{d2}(n)$ are merely two variations to the same theme. However, whether or not the logarithm is used can play an important role in the asymptotic behavior of the tests as well as in their small sample performance.

Essentially, (7) and (8) try to detect departures from the equality $\mathbf{B}(\boldsymbol{\xi}_0) = -\mathbf{A}(\boldsymbol{\xi}_0)$ using the determinant of the matrix $\mathbf{B}_n(\boldsymbol{\xi})[-\mathbf{A}_n^{-1}(\boldsymbol{\xi})]$. The use of the determinant is a plausible and sensible choice to quantify the “distance” between $\mathbf{B}_n(\boldsymbol{\xi})$ and $-\mathbf{A}_n^{-1}(\boldsymbol{\xi})$. Another intuitive and appealing possibility consists in combining the determinant with the trace into the same test statistic to quantify the “distance” between the two matrices of interest. We will explore this approach in the next section.

3.2 The Determinant-Trace Test

In this section, both the trace and the determinant of $\mathbf{A}_n(\boldsymbol{\xi}_0)$ and $\mathbf{B}_n(\boldsymbol{\xi}_0)$ are incorporated in the test statistic. Define

$$\delta_{dt}(n) = \frac{\text{tr}[\mathbf{B}_n(\boldsymbol{\xi}_0)]}{\text{tr}[-\mathbf{A}_n(\boldsymbol{\xi}_0)]} - \frac{|\mathbf{B}_n(\boldsymbol{\xi}_0)|}{|-\mathbf{A}_n(\boldsymbol{\xi}_0)|}, \quad (9)$$

The following result allows us to establish the distribution of $\delta_{dt}(n)$ under the null hypothesis.

Theorem 2 *If the model is correctly specified then under general regularity conditions*

$$\frac{n[\delta_{dt}(n)]^2}{2\sigma_{\delta_n}} \sim \chi_1^2. \quad (10)$$

Like in the previous section the fact that $\mathbf{A}_n(\boldsymbol{\xi}_0)$ is a consistent estimator of $\mathbf{A}(\boldsymbol{\xi}_0)$ implies that the asymptotic distribution of $\delta_{dt}(n)$ is the same as the asymptotic distribution of

$$\delta_{dt}^*(n) = \sum_{k=1}^p \left(\frac{\hat{\gamma}_{kn}}{\sum_l \gamma_l} \right) - \prod_{k=1}^p \frac{\hat{\gamma}_{kn}}{\gamma_k}. \quad (11)$$

If we further denote

$$\sigma_{\delta_n} = \sum_{k=1}^p \left(\frac{\gamma_k}{\sum_l \gamma_l} - 1 \right)^2,$$

the distribution of $\delta_{dt}^*(n)$ follows from applying the delta method. Indeed, the gradient of $\delta_{dt}^*(n)$, evaluated in $\hat{\gamma}_n = \boldsymbol{\gamma}$, corresponds to

$$\left. \frac{\partial \delta_{dt}^*(n)}{\partial \hat{\gamma}_{kn}} \right|_{\hat{\gamma}_n = \boldsymbol{\gamma}} = \frac{1}{\sum_l \gamma_l} - \left(\prod_{\ell \neq k} \frac{\hat{\gamma}_{\ell n}}{\gamma_\ell} \right) \frac{1}{\gamma_k} \Big|_{\hat{\gamma}_n = \boldsymbol{\gamma}} = \frac{1}{\sum_l \gamma_l} - \frac{1}{\gamma_k}.$$

If we let $\boldsymbol{\Delta}^T = \left(\frac{1}{\sum_l \gamma_l} - \frac{1}{\gamma_1} \dots \frac{1}{\sum_l \gamma_l} - \frac{1}{\gamma_p} \right)$, then $\delta_{dt}^*(n) \sim N\left(0, \frac{2}{n} \boldsymbol{\Delta}^T \boldsymbol{\Gamma}^2 \boldsymbol{\Delta}\right)$. Since $\boldsymbol{\Delta}^T \boldsymbol{\Gamma}^2 \boldsymbol{\Delta} = \sigma_{\delta_n}$, this results in (10).

In principle, other alternatives using these eigenvalues could be considered as well. We have chosen three that are intuitively appealing and mathematically tractable when calculating their null distribution. Note further that these tests are based on some assumptions. For instance, the assumption that the individual contributions to the score are normally distributed. Obviously, departures from this assumption may affect the distributional results in (6) and the performance of the proposed tests. In this case, [Waternaux \(1976\)](#) showed that the estimators of the eigenvalues obtained from the observed covariance

matrix will still be normally distributed and centered around their population values, however the covariance matrix of these estimators may require a correction depending on the shape of the real distribution. Another important issue is the substitution of $\boldsymbol{\xi}_0$ by its consistent estimator under the null, $\widehat{\boldsymbol{\xi}}_n$. This extra variability introduced by using an estimator is not explicitly taken into account by the tests. Essentially, these assumptions are the price to pay in order to gain simplicity and avoid the use of high order derivatives.

In what follows, we will empirically study the performance of these tests via simulations. This study will help us to evaluate the impact of these assumptions and the behavior of these asymptotic results in finite sample sizes.

3.3 Implementation

One major advantage of the diagnostic tools proposed in this manuscript is their relatively easy implementation using standard software, like the SAS procedures NLMIXED and IML. Indeed, note first that the Hessian $\mathbf{A}_n(\widehat{\boldsymbol{\xi}}_n)$ follows directly from NLMIXED. The subject's contribution to $\mathbf{B}_n(\widehat{\boldsymbol{\xi}}_n)$ can also be obtained in a relative straightforward way, but they need some extra calculations. To compute these values, we need to fit the final model in NLMIXED to each subject separately, keeping the maximum likelihood estimates fixed by setting `maxiter = 0`, and saving the corresponding first order derivatives. This approach is illustrated with more detail using some exemplary SAS code in Appendix A.

It is also important to point out that the maximum likelihood estimates can be subject to bias due to rounding errors and inadequate precision of the sta-

tistical software. These numerical issues have been studied, for example, by [Lesaffre and Spiessens \(2001\)](#). They have shown that the default number of quadrature points calculated by NLMIXED can lead to inaccurate maximum likelihood estimates and p -values. In this manuscript, all analyses were carried out using adaptive Gaussian quadrature and 50 quadrature points to approximate the likelihood function. Note further that the previously introduced tests involved the first and second order derivatives of the likelihood function, rendering the approximations more challenging. It is difficult to establish a general rule to define the number of quadrature points necessary to achieve a reasonable approximation in this case. However, our empirical experience with these and previous simulations indicate that 50 are generally sufficient to obtain adequate approximations to the likelihood function and its derivatives. In general, it would always be advisable not to restrict the computations to the default number of quadrature points given by NLMIXED and explore other values. For instance, one could sequentially increase the number of quadrature points until the results stabilize and no relevant differences are observed with further increases. Using 50 quadrature points to estimate $\mathbf{A}_n(\hat{\boldsymbol{\xi}}_n)$ and $\mathbf{B}_n(\hat{\boldsymbol{\xi}}_n)$, we also obtained stable results, when using the SAS procedure IML, to determine the eigenvalues of these two matrices.

4 Simulations

Although misspecification of the random effects distribution can severely affect the estimation and inferential procedures in GLMM, other types of misspecifications of the random effects structure are important as well. Indeed, as stated in the introduction, simulations by [Heagerty and Kurland \(2001\)](#) have shown

that misspecifications such as assuming a logistic-normal model, (i) when the variance of the random effect depends on a covariate in the mean structure, (ii) when the random structure includes both a random intercept and slope, or (iii) when the random effects are auto-correlated, can also induce an important bias in the estimates of the fixed effects parameters. Therefore, in this section, we will also evaluate the performance of the proposed tests to detect these more general misspecifications of the random-effects structure.

4.1 *Non-normal Random Effects*

Using simulations with the model given by (4), while considering different distributions for the random intercept, [Litière et al. \(2008\)](#) established that the variance component estimator is always subject to considerable bias when the random-effect distribution is misspecified. Further, they found that bias can also be present in the estimates of the mean structure parameters, depending on the size of the variance associated with the random intercept. In this section, we will use a subset of these simulations to evaluate the power of the new proposals to detect this type of misspecification. The binary responses were generated using model (4), including an intercept, a binary covariate z_i , a within-cluster covariate t_j taking values 0, 1, 2, 4, 6, and 8, and a random intercept b_{0i} sampled from five distinct mean-zero distributions, each with variances $\sigma_b^2 = 4$ and 32: a normal, a power-function, and a lognormal distribution, as well as a discrete distribution with equal probability at two support points, and an asymmetric mixture of two normal densities. Note that $\sigma_b^2 = 32$ allows to investigate scenarios with variances in the same order of magnitude as the one observed in the case study, whereas $\sigma_b^2 = 4$ is used to analyze the

performance of the tests in less extreme scenarios.

The parameters in the mean structure were fixed at $\beta_0 = -8$, $\beta_1 = 2$, and $\beta_2 = 1$, in accordance with the values estimated from the case study. Six different sample sizes were considered, namely 50, 100, 200, 350, 500, and 1000 subjects, and for each setting, 500 data sets were generated. The diagnostic tools were then used to assess the appropriateness of model (4), with a normal random intercept, for the analysis of each of the generated data sets. For each test, we determined the proportion of cases in which a significant result, at a 5% level, was detected. When the random effects were generated from a normal distribution, this proportion corresponds to the type I error rate; in the other settings, it represents the power of the tests to detect the misspecification. The findings of these simulations are shown in Table 2.

The first part of the table displays the results for the small variance setting ($\sigma_b^2 = 4$). In this scenario, most of the tests exhibit a reasonable type I error rate for samples of 200 or 350 subjects. The only exception seems to be the determinant-trace test, showing a considerably inflated type I error rate of up to 16%. The determinant tests have the best global behavior when both power and type I error rate are taken into account. Especially the test based on $\delta_{d1}(n)$ successfully detects the discrete and lognormal distribution for sample sizes of 350 or larger. The determinant-trace test also produced good power values but its inflated type I error rate would make its results unclear in a real situation where the true distribution is unknown.

The second part of the table summarizes the results for the large variance scenario ($\sigma_b^2 = 32$). Unlike in the small variance setting, here all the tests exhibit a very good type I error rate, even for small samples of 50 subjects.

Table 2

Power and type I error rate of the determinant tests $\delta_{d1}(n)$ and $\delta_{d2}(n)$, and the determinant-trace test $\delta_{dt}(n)$ to detect a misspecified random-effects distribution: a normal random intercept is assumed, whereas the random effect was generated from a normal (No), a power-function (PF), a discrete (D), an asymmetric mixture of two normals (AM) or a lognormal distribution (LN), each with variance $\sigma_b^2 = 4$ or 32.

		$\sigma_b^2 = 4$			$\sigma_b^2 = 32$		
	n	$\delta_{d1}(n)$	$\delta_{d2}(n)$	$\delta_{dt}(n)$	$\delta_{d1}(n)$	$\delta_{d2}(n)$	$\delta_{dt}(n)$
No	50	0.126	0.010	0.048	0.074	0.042	0.032
	100	0.106	0.030	0.098	0.052	0.026	0.020
	200	0.096	0.054	0.126	0.048	0.044	0.032
	350	0.098	0.072	0.152	0.062	0.050	0.026
	500	0.100	0.084	0.164	0.054	0.040	0.018
	1000	0.074	0.068	0.126	0.052	0.038	0.012
PF	50	0.167	0.004	0.094	0.422	0.014	0.448
	100	0.208	0.034	0.240	0.542	0.214	0.754
	200	0.262	0.092	0.422	0.734	0.558	0.968
	350	0.334	0.200	0.570	0.926	0.846	0.996
	500	0.392	0.278	0.648	0.990	0.980	1.000
	1000	0.640	0.580	0.894	1.000	1.000	1.000
D	50	0.412	0.010	0.266	0.914	0.458	0.840
	100	0.502	0.176	0.600	0.986	0.942	0.992
	200	0.716	0.550	0.858	1.000	0.998	1.000
	350	0.894	0.782	0.972	1.000	1.000	1.000
	500	0.954	0.926	0.994	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000
AM	50	0.145	0.006	0.095	0.266	0.014	0.500
	100	0.116	0.016	0.148	0.322	0.094	0.722
	200	0.152	0.054	0.288	0.456	0.264	0.904
	350	0.164	0.086	0.328	0.602	0.474	0.998
	500	0.170	0.108	0.404	0.704	0.602	1.000
	1000	0.312	0.260	0.620	0.934	0.918	1.000
LN	50	0.076	0.088	0.169	0.091	0.113	0.178
	100	0.267	0.369	0.583	0.172	0.260	0.474
	200	0.676	0.766	0.936	0.476	0.578	0.868
	350	0.914	0.954	0.996	0.690	0.786	0.982
	500	0.994	0.996	1.000	0.874	0.924	1.000
	1000	1.000	1.000	1.000	0.998	1.000	1.000

Regarding power, all tests perform quite well already for sample sizes of 200 or larger. The best overall performance is now observed for the determinant-trace test. Indeed, it shows a very good type I error rate and a remarkable power for all misspecifications studied in this scenario.

Clearly, all tests perform considerably better when the variance of the random effect is large. This is a desirable behavior given the results obtained by [Litière et al. \(2008\)](#). As stated before, these authors found that considerable bias could appear under misspecification when the variance was large. Precisely, the scenario where the proposed tests showed the larger power and best general performance.

4.2 *Random Intercept Variance Depending on a Binary Covariate*

Similar to the simulation approach followed by [Heagerty and Kurland \(2001\)](#), binary responses were now generated using the model

$$\text{logit}\{P(y_{ij} = 1|b_{ij})\} = \beta_0 + \beta_1 z_i + \beta_2 t_j + \beta_3 z_i t_j + b_{ij}, \quad (12)$$

where z_i is a binary covariate, t_j is a within-cluster covariate representing a linear trend, with $t_j = (j - 1)/(n_i - 1)$, the variance of the random intercept $b_{ij} = b_{i0}$ is sampled from a distribution given by

$$b_{i0} \sim \begin{cases} N(0, \sigma_0^2) & \text{when } z_i = 0 \\ N(0, \sigma_1^2) & \text{when } z_i = 1, \end{cases} \quad (13)$$

and $n_i = 6$. The parameters in the mean structure were fixed at $\beta_0 = -2$, $\beta_1 = 1$, $\beta_2 = 0.5$ and $\beta_3 = -0.25$. In total, 500 data sets, each containing information on $n = 500$ subjects, were generated using the previous spec-

ifications, and model (12) was fitted to the generated data, assuming that $b_{ij} = b_{i0} \sim N(0, \sigma_b^2)$. In this setting, Heagerty and Kurland (2001) found that substantial bias can occur for all coefficients when σ_0 and σ_1 are very different. For example, they reported 38% and 31% of relative bias in the estimation of β_1 and β_3 , respectively, when $\sigma_0 = 1$ and $\sigma_1 = 2$. Additionally, they observed that, as the discrepancy between the two parameters increases, so does the bias in the parameter estimates.

To study the performance of our proposals in this particular scenario, we applied the tests to the generated data sets and determined the proportion out of the 500 repetitions in which the tests were able to detect the misspecification (at a 5% significance level). The corresponding powers are displayed in the first panel of Table 3 as a function of σ_0 and σ_1 . Note that, when $\sigma_0 = \sigma_1$, these values correspond to the type I error rate.

Remarkably, all the tests have in general a poor performance in this setting. The observed type I error rate exceeds the pre-specified value in some scenarios and the power is usually very small. They failed to detect the misspecification, even when the difference between σ_0 and σ_1 was largest. For instance, the determinant-trace test $\delta_{dt}(n)$ shows an excellent power when $\sigma_1 = 3.0$ and $\sigma_0 = 0.5$, but fails to detect the reverse situation, when $\sigma_1 = 0.5$ was combined with $\sigma_0 = 3.0$, in 75% of the cases.

4.3 Ignoring a Random Effect

Another type of misspecification in the random structure occurs when a random slope is incorrectly ignored. To study the performance of our proposals

Table 3

Power of the determinant tests $\delta_{d1}(n)$ and $\delta_{d2}(n)$, and the determinant-trace test $\delta_{dt}(n)$ to detect model misspecification, when a logistic-normal model is assumed, but (a) the variance of the random intercept depends on a binary cluster-level covariate, $[b_{i0}|z_i = 0] \sim N(0, \sigma_0^2)$ and $[b_{i0}|z_i = 1] \sim N(0, \sigma_1^2)$, (b) the data are generated using both a random intercept and slope ($b_{ij} = b_{i0} + b_{i1}t_j$), with variance σ_0^2 and σ_1^2 , respectively, and (c) the data are generated using autocorrelated random effects b_{ij} such that $\text{cov}(b_{ij}, b_{ik}) = \sigma^2 \rho^{|t_{ij} - t_{ik}|}$.

(a)					(b)					(c)				
σ_1	σ_0	$\delta_{d1}(n)$	$\delta_{d2}(n)$	$\delta_{dt}(n)$	σ_1	σ_0	$\delta_{d1}(n)$	$\delta_{d2}(n)$	$\delta_{dt}(n)$	ρ	σ	$\delta_{d1}(n)$	$\delta_{d2}(n)$	$\delta_{dt}(n)$
0.5	0.5	0.100	0.080	0.106	0.2	0.5	0.078	0.070	0.110	0.5	0.5	0.067	0.080	0.103
	1.0	0.056	0.060	0.090		1.0	0.056	0.040	0.076		1.0	0.182	0.267	0.333
	2.0	0.176	0.108	0.026		2.0	0.050	0.048	0.048		2.0	0.852	0.910	0.938
	3.0	0.594	0.478	0.248		3.0	0.070	0.060	0.042		3.0	0.992	0.998	0.998
1.0	0.5	0.134	0.072	0.184	0.5	0.5	0.070	0.068	0.094	0.7	0.5	0.095	0.099	0.128
	1.0	0.050	0.042	0.070		1.0	0.066	0.068	0.102		1.0	0.232	0.306	0.366
	2.0	0.072	0.048	0.044		2.0	0.062	0.090	0.086		2.0	0.954	0.970	0.972
	3.0	0.227	0.165	0.107		3.0	0.058	0.067	0.053		3.0	1.000	1.000	1.000
2.0	0.5	0.364	0.236	0.638	0.8	0.5	0.068	0.116	0.120	0.9	0.5	0.076	0.062	0.103
	1.0	0.154	0.116	0.242		1.0	0.120	0.166	0.172		1.0	0.092	0.148	0.190
	2.0	0.062	0.046	0.064		2.0	0.178	0.242	0.224		2.0	0.768	0.864	0.834
	3.0	0.062	0.044	0.028		3.0	0.190	0.234	0.198		3.0	0.998	0.998	0.998
3.0	0.5	0.740	0.584	0.926	1.0	0.5	0.138	0.178	0.174					
	1.0	0.444	0.328	0.654		1.0	0.210	0.300	0.260					
	2.0	0.100	0.068	0.086		2.0	0.356	0.434	0.380					
	3.0	0.066	0.064	0.046		3.0	0.368	0.468	0.394					

in this scenario, we have generated binary responses from the model given by (12), with $b_{ij} = b_{i0} + b_{i1}t_j$, and σ_0^2 and σ_1^2 representing the variance of the random intercept b_{i0} and slope b_{i1} , respectively.

Simulations by [Heagerty and Kurland \(2001\)](#) showed that when these data are analyzed wrongly assuming that $b_{ij} = b_{i0}$, moderate bias can appear in the estimation of the regression coefficients. For instance, they observed asymp-

otic relative biases as large as 30-50% in the estimates of β_2 and β_3 when σ_0 is small and σ_1 is large. On the other hand, the bias for the estimators of the intercept β_0 and the cluster-level covariate effect β_1 remained below 15% for all considered pairs of (σ_0, σ_1) .

The second panel of Table 3 shows the power of the diagnostic tools to detect this type of misspecification, as a function of σ_0 and σ_1 . As one would expect, all tests fail to detect the misspecification when σ_1 is small. However, the results of Heagerty and Kurland (2001) showed that little bias is present in this case. When σ_1 is increased, relative to σ_0 , then the power slightly improves. Nevertheless, when $\sigma_1 = 1$ and $\sigma_0 = 0.5$, precisely the setting in which bias as large as 52% was obtained for β_2 and β_3 , we only observed a power of at the most 17% with the determinant-trace test.

4.4 Autoregressive Random Effects

In the analysis of longitudinal data one often observes that the dependence between repeated measurements within a subject seems to decay as the time separation between the measurements increases. This could be accounted for with a generalized linear mixed model including autocorrelated random effects b_{ij} for which $\text{cov}(b_{ij}, b_{ik}) = \sigma^2 \rho^{|t_{ij} - t_{ik}|}$. Simulations by Heagerty and Kurland (2001) with this type of misspecification have shown that substantial negative bias can arise in the estimated fixed effects, with increasing bias as σ increases, and especially when ρ is small. Note that the random intercept model follows as a special case of the autoregressive model when $\rho = 1$. For models with $\rho < 1$, a potentially large negative bias can be observed in $\hat{\sigma}_n$, given that it estimates the common variance and therefore approximates the true covariances

$\sigma^2 \rho^{|t_{ij}-t_{ik}|}$. These authors observed that as ρ decreases, the negative bias in $\hat{\sigma}_n$ increases, ranging between -30% and -50% when $\rho = 0.7$, and between -47% and -70% when $\rho = 0.5$. As a result, substantial negative bias can also arise in the estimated regression coefficients, with increasing bias as σ increases. For instance, when $(\rho, \sigma) = (0.5, 3.0)$ negative bias as high as -45% occurred in each of the mean structure parameter estimates.

The final panel of Table 3 shows the power of the diagnostic tools to detect this type of misspecification, as a function of σ and ρ . From the table it follows that, unlike in the previous misspecification settings, as of $\sigma \geq 2$, the tools are able to detect the misspecification in over 80% of the cases. Given that the bias in the estimation of the mean structure parameters was seen to be more substantial as of $\sigma \geq 2$ (Heagerty and Kurland, 2001), this is a very desirable behavior. The results obtained with this misspecification are specially appealing if we take into account that the test introduced by Tchetgen and Coull (2006) will not be applicable in this scenario.

5 Analysis of Case Study

In this section, we will apply the different tests to assess the suitability of model (4) with normal random intercept, for the analysis of the case study.

The determinant tests lead to $\delta_{d1}(n) = 0.075$ and $\delta_{d2}(n) = 1.078$ with corresponding p -values of 0.763 and 0.754, respectively. On the other hand, the determinant-trace test delivers the following output: $n[\delta_{dt}(n)]^2/2\sigma_{\delta_n} = 0.001$ accompanied by $p = 0.996$.

Even though these results suggest that the data at hand do not provide evi-

dence for misspecification, one should interpret them with due caution. Note first that the simulations clearly showed that the tests may lack power to detect misspecifications in the presence of small samples sizes, such as the one considered here ($n = 128$). Additionally, the simulations also illustrated that these diagnostic tools can fail to detect some specific types of misspecifications even when relative large sample sizes are available. Therefore, they should only be considered as an extra element in the evaluation of the model and not as a definitive instrument. In this spirit, an additional sensitivity analysis was also conducted, where different distributions were considered for the random intercept (Litière et al. (2008)). The estimates of the parameters of interest and the associated inferential procedures were similar, irrespective of the distribution used to obtain them.

Finally, the results of this sensitivity analysis, together with the wide variety of models considered in the model building exercise and the clearly non-significant p -values obtained from the diagnostic tests, offer a comfortable level of confidence about the conclusions emanating from the analysis based on model (4).

6 Discussion

It is known that the maximum likelihood estimators and the associated inferential procedures can be affected by misspecifications of the random-effects structure in GLMM. In this manuscript, we have provided a toolbox of tests to detect misspecifications of this type. The proposals considered have the strong advantage of not requiring higher order derivatives of the likelihood function, an important issue in situations where we have to deal with a complex likeli-

hood without closed form. We compared, via simulations, the performance of these tests. Encouragingly, the simulations have shown a reasonable behavior, especially in some scenarios where the cost of a misspecified random-effects distribution is higher.

Clearly, the tests did not exhibit the same power in all the settings considered, however, this is a frequently encountered situation. Indeed, inferential procedures are usually more powerful in some regions of the parametric space, defined by the alternative hypothesis, than others. Moreover, the lack of a clear winner among the tests considered should not be considered a disappointing result. Uniformly most powerful tests (UMPT) are already difficult to obtain in simpler scenarios where the alternative hypothesis is just a subset of the p -dimensional space. Here the alternative is defined by a set of distribution functions, making the chances of finding a UMPT, formally or informally, even smaller.

Notice that these diagnostic tools are, in principle, suitable to detect different types of model misspecifications. In this manuscript, we have explored their performance when facing misspecifications in the random-effects structure. Supplementally, simulations not shown here using a logistic random-intercept model with a misspecified mean, for example, by omitting an important covariate, but correctly specified random-effects structure, showed that the effect of the covariate has to be considerably large for the tests to detect the misspecification. This is an expectable result given that random effects are known to account for heterogeneity caused by omitting certain explanatory variables (Agresti, 2002). In this case, the tests are not nearly as powerful as, for instance, a Wald test, which explicitly tests for the presence of a covariate in the model. Still, one should not forget that the alternative hypothesis

of the Wald test consists of $H_1 : \beta_1^0 \neq 0$, whereas the alternative hypothesis of the proposed diagnostic tools consists of a wide range of model misspecifications. Other types of misspecification are also possible, for instance, one could misspecify the link function or not take into account the presence of overdispersion, etcetera.

Finally, it is important to point out that a non-significant result from these diagnostic tools does not necessarily imply that there is no misspecification. For instance, as can be seen from columns (a) and (b) in Table 3, due to a lack of power a non-significant result can still mask some types of (random-effects) misspecification. On the other hand, a significant result will not necessarily imply that there is a problem in the random structure of the model. Nevertheless, given that the tools seem to be especially powerful in detecting misspecification of the random-effects distribution, this could be one of the aspects to check if such a significant result is obtained.

Acknowledgment

Financial support from the IAP research network #P6/03 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

Appendix A. SAS Code

First, we run the statistical model to obtain the maximum likelihood estimates and the Hessian. For, instance, consider model (4) for the analysis of the schizophrenia data.

```

proc nlmixed data=new qpoints=50 hess;
  parms beta0=1 beta1=1 beta2=1 s2u1=1;
  eta = beta0 + beta1*Z + beta2*time + u;
  expeta=exp(eta);
  p=expeta/(1+expeta);
  model Y ~ binary(p);
  random u ~ normal(0,s2u1) subject=subject;
  ods output ParameterEstimates=pe Hessian=hess;
run;

```

Next, we run the following macro, which will determine each subjects contribution to $\mathbf{B}_n(\boldsymbol{\xi})$. Note that the subjects in the data require an identification number, going from 1 to n , the total number of subjects.

```

%macro individual(data=, n=);
data glmmgrad; set _NULL_; run;
%do i = 1 %to &n;
data hulp;
  set &data;
  if subject ^= &i then delete;
run;
proc nlmixed data=hulp qpoints=50 maxiter=0;
parms / data=pe;
  eta = beta0 + beta1*Z + beta2*time + u;
  expeta=exp(eta);
  p=expeta/(1+expeta);
  model Y ~ binary(p);
  random u ~ normal(0,s2u1) subject=subject;

```

```

ods output ParameterEstimates=pe1;
run;
data glmmgrad; set glmmgrad pe1; run;
%end;
%mend;
%individual(data=new, n=128);

```

Now the file glmmgrad contains first order derivatives of the likelihood for each subject in the data set. Finally, we use IML to calculate the test statistics.

```

proc iml; reset noprint;
n=128;

/* determining A */
use hess; read all into hess;
na = ncol(hess);
np = nrow(hess);
A = -hess[,2:na]; An = A/n;

/* determining B */
use glmmgrad; read all into gradient;
i=1; B = {0};
begin=1;
do while (i <= n);
    end = begin -1 + np;
    first = gradient[begin:end, 2];
    B = B + first*first';
    begin = end + 1;
end;

```

```

        i = i+1;
end;
Bn = B/n;

/* the determinant tests */
delta1 = log(det(-Bn*inv(An)));
test_det1 = (n/(2*np))*(delta1**2);
prob_det1 = 1 - PROBCHI(test_det1,1);
delta2 = det(Bn)*det(-inv(An));
test_det2 = (n/(2*np))*((delta2-1)**2);
prob_det2 = 1 - PROBCHI(test_det2,1);

/* the determinant-trace tests */
delta_dt = trace(Bn)/trace(-An) - det(Bn)/det(-An);
Eb = eigval(-An);
sigma = sum((Eb/trace(-An) - 1)**2)
test_dettr = (n/2)*(delta_dt**2)/sigma;
prob_dettr = 1-probchi(test_dettr,1);
quit; run;

```

References

- Agresti, A., 2002. Categorical data analysis. Wiley, Hoboken, N.J.
- Agresti, A., Caffo, B. and Ohman-Strickland, P., 2004. Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Comput. Statist. Data Anal.* 47, 639–653.
- Alonso, A., Geys, H., Molenberghs, G., Kenward, M.G. and Vangeneugden, T.,

2004. Validation of surrogate markers in multiple randomized clinical trials with repeated measurements: canonical correlation approach. *Biometrics* 60, 845–853.
- Anderson, T.W., 1963. Asymptotic theory for principal component analysis. *Ann. Math. Stat.* 34, 122–148.
- Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. *J. Am. Stat. Ass.* 88, 9–25.
- Chen, J., Zhang, D. and Davidian, M., 2002. A Monte Carlo EM algorithm for generalized linear mixed models with flexible random-effects distribution. *Biostatistics* 3, 347–360.
- Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger S.L., 2002. *Analysis of Longitudinal Data*. Oxford University Press, New York.
- Girschick, M.S., 1939. On the sample theory of the roots of determinantal equations. *Ann. Math. Stat.* 10, 203–224.
- Heagerty, P.J. and Kurland, B.F., 2001. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 88, 973-985.
- Kenward, M.G. and Molenberghs, G., 2007. *Missing Data in Clinical Studies*. Wiley, Chichester.
- Lesaffre, E. and Spiessens, B., 2001. On the effect of the number of quadrature points in a logistic random-effects model: an example. *Appl. Stat.* 50, 325-335.
- Lehmann, E.L, 1999. *Elements of Large Sample Theory*. Springer-Verlag, New York, Inc.
- Litière, S., Alonso, A. and Molenberghs, G., 2007. Type I and type II error random-effects misspecification in generalized linear mixed models. *Biometrics* 63, 1038–1044.
- Litière, S., Alonso, A. and Molenberghs, G., 2008. The impact of a misspecified

- random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine* 00, 000–000.
- Molenberghs, G. and Verbeke, G., 2005. *Models for Discrete Longitudinal Data*. Springer, New York.
- Neuhaus, J.M., Hauck, W.W. and Kalbfleisch, J.D., 1992. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* 79, 755–762.
- Tchetgen, E.J. and Coull, B.A., 2006. A diagnostic test for the mixing distribution in a generalized linear mixed model. *Biometrika* 93, 1003–1010.
- Waagepetersen, R., 2006. A simulation-based goodness-of-fit test for random effects in generalized linear mixed models. *Scand. J. Stat.* 33, 721–731.
- Waternaux, C.M., 1976. Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika* 63, 639–645.
- White, H., 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-25.
- Wolfinger, R. and O’Connell, M., 1993. Generalized linear mixed models: a pseudo-likelihood approach. *J. Statist. Comput. Simulation* 48, 233–243.