T E C H N I C A L
R E P O R T

0658

A SANDWICH-ESTIMATOR TEST FOR MISSPECIFICATION IN MIXED-EFFECTS MODELS

LITIERE S., ALONSO A., and G. MOLENBERGHS

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

A Sandwich-Estimator Test for Misspecification in Mixed-Effects Models

Saskia Litière*, Ariel Alonso, and Geert Molenberghs

Center for Statistics, Hasselt University, Agoralaan, Building D, B3590 Diepenbeek, Belgium

E-mail: saskia.litiere@uhasselt.be

Author's Footnote

Geert Molenberghs is Professor of Biostatistics, Ariel Alonso is Postdoctoral Researcher, and Saskia Litière is Research Assistant, Center for Statistics, Hasselt University, Diepenbeek, Belgium (email: saskia.litiere@uhasselt.be). The authors gratefully acknowledge support from Belgian IUAP/PAI network "Statistical Techniques and Modeling for Complex Substantive Questions with Complex Data".

Abstract

Generalized linear mixed models have become a frequently used tool for the analysis of non-Gaussian longitudinal data. Estimation is often based on maximum likelihood theory, which assumes that the underlying probability model is correctly specified. Recent research is showing that the results obtained from these models are not always resistant against departures from the assumptions on which these models are based. Therefore, diagnostic tools for the detection of model misspecifications are of the utmost importance. In this paper, we propose a diagnostic test that provides an alternative to the Information-Matrix Test (IMT) proposed by White (1982). However, unlike the IMT, the new test does not involve third order partial derivatives of the log-likelihood, making it easy implementation-wise and diminishing its computational complexity. We study the power and type I error rate of the new test to detect random-effects misspecification via simulations. In the context of linear mixed models, the new proposal performs as good as the IMT. The new test enjoys good performance for generalized linear mixed models, which is important since the IMT is very difficult to apply in this context, owing to the lack of a closed form for the likelihood.

Keywords: Linear mixed model; Generalized linear mixed model; Information-matrix test; Power; Type I error.

1. INTRODUCTION

Over the last fifteen years, hierarchical models have become a frequently used tool in statistical analysis. Their implementation in popular statistical packages like, e.g., the SAS procedures MIXED, NLMIXED, and GLIMMIX, or the R functions `lme` and `nlme`, have substantially contributed to their wide spread. Often, the model parameters are estimated using maximum likelihood (ML), hence resting on the assumption that the model is correctly specified. Therefore, violations of some model assumption can affect our inferences. One of the basic assumptions underlying hierarchical models concerns the distribution of the random effects. Also to improve mathematical tractability and numerical performance, random effects are commonly assumed to be normally distributed. Unfortunately, since random effects are not observed, the validity of this assumption is difficult to verify. So, naturally, we are interested in the impact of misspecifying the random-effects distribution on the maximum likelihood estimators and other inferential procedures, as well as in the development of diagnostic tools that can help us to detect this type of misspecification.

In linear mixed models (LMM), Verbeke and Lesaffre (1997) showed that the maximum likelihood estimators for fixed effects and variance components, obtained under the assumption of normal random effects, are consistent and asymptotically normally distributed, even when the random-effects distribution is not normal. Nevertheless, recent research suggests that this does not hold for generalized linear mixed models (GLMM). For example, Neuhaus, Hauck, and Kalbfleisch (1992) showed that the maximum likelihood estimators of a random-intercept logistic model with misspecified random-effects distribution are inconsistent, but that the magnitude of the bias is typically small. Simulations by Chen, Zhang, and Davidian (2002) also indicate that the estimation of the regression coefficients may be subject to negligible bias only, under misspecification of the random-effects distribution. According to Agresti, Caffo, and Ohman-Strickland (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, e.g., by omitting an influential binary covariate, this can affect the predictive qualities of characteristics involving the random effects as well as the fixed effects. Litière *et al.* (2006) found, using simulations with a random-intercept logistic model and a wide range of distributions for the random effect, that the estimates of the variance components are always subject to considerable bias when the random-effects distribution is misspecified. Litière, Alonso, and Molenberghs (2007) established that type I error and power can be severly impacted when misspecifying the random-effects distribution. On the other hand, the bias induced in the estimates of the mean structure parameters appears to depend on the magnitude of the variance components, whereby large bias is associated with large random-effects variances. Clearly, in a practical situation, the bias present in the variance component estimators, under misspecification, will make it hard to distinguish between the two scenarios, i.e., small or larger variance components. Therefore, it can be difficult to determine how severe the impact on the mean parameters can be.

Note that the severe bias observed in the variance components can have an important impact in studies where the variance components are of main interest. This is the case, for instance, in fields like surrogate marker validation, reliability of rating scales, or studies of the criterion and predictive validity of psychiatric scales.

Evidently, in these circumstances, the development of diagnostic tools is of great importance. White (1982) already proposed a general test for model misspecification. Unfortunately, his 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 in LMM, it can become an important problem when such a closed form does not exist like, for example, in GLMM.

In the present work, we propose an alternative diagnostic tool along the ideas of the Information-Matrix Test, introduced by White (1982), but without the need for third-order partial derivatives of the likelihood. The motivating case study will be introduced in Section 2. A brief overview of the Information-Matrix Test is given in Section 3, followed by the introduction of the new test in Section 4. Next, in Section 5, the performance of our proposal is evaluated via simulations and under both LMM and GLMM. Finally, the case study is analyzed in Section 6.

2. CASE STUDY

The case study consists of individual patient data from a randomized clinical trial, comparing the effect of risperidone to conventional antipsychotic agents for the treatment of chronic schizophrenia (Alonso *et al.* 2004). Several measures can be used to assess a patient's global condition. The *Clinical Global Impression* (CGI) is generally accepted as a subjective clinical measure of change. It is a 7-grade scale used to characterize a subject's mental condition. Our binary response variable Y is a dichotomous version of this scale which equals 1 for patients classified as normal to mildly ill, and 0 for patients classified as moderately to severely ill. Since it has been established that risperidone is most effective at doses ranging from 4 to 6 mg/day, we included only those patients receiving either these doses of risperidone, i.e., the treatment group $z_i = 1$, or an active control, i.e., the control group $z_i = 0$ for patient $i = 1, \ldots, N$. Treatment was administered for 8 weeks and the outcome was measured at 6 fixed time points $t_{ij} = t_j$, taking values 0, 1, 2, 4, 6, and 8 weeks and corresponding to occasion indicator $j = 1, \ldots, n_i = n = 6$. One hundred twenty-eight patients were included into the trial, implying $N = 128$.

Previous data analysis has shown that an adequate model for these data is given by (Litière *et al.* 2006):

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

where i and j denote the patient and the measurement occasion, respectively, and b_{0i} is a random effect, assumed to follow a normal distribution. The maximum likelihood estimates of the fixed effects are given in Table 1, and the variance of the random effect was estimated as 21.01 (s.e. 6.81). Note that a large random-effects variance such as this one is thus not rare in clinical trials where, for example little variability in the response is expected in the placebo group and considerably more variability in the treated group. In such a scenario, the mean parameters, including the treatment-effect parameters, could be subject to serious bias under misspecification. Given this and the discussion in the previous section, one would like to test for possible misspecification of the random-effects distribution in particular, or any other possible misspecification of the model in general. In the following sections, we will describe two such tools.

3. THE INFORMATION-MATRIX TEST

Let us consider a random variable $\boldsymbol{y}$ with density function g, and a parametric family of density functions $\mathfrak{F} = \{f(\boldsymbol{y};\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Gamma\}$. If there exists a $\boldsymbol{\xi}_0 \in \Gamma$ such that $g(\boldsymbol{y}) = f(\boldsymbol{y},\boldsymbol{\xi}_0)$, then the maximum likelihood estimator $\hat{\boldsymbol{\xi}}_n$ of $\boldsymbol{\xi}_0$ is consistent and asymptotically normal. However, since in practice g is unknown, it is difficult to check whether g belongs to $\mathfrak{F}$ or not. Fortunately, White (1982) found that, under general regularity conditions, the maximum likelihood estimator $\hat{\boldsymbol{\xi}}_n$ will (strongly) converge to the value of $\boldsymbol{\xi}$, denoted by $\boldsymbol{\xi}^*$, which minimizes the so-called Kullback-Leibler Information Criterion (KLIC):

$$I(g : f,\boldsymbol{\xi}) = E\Big\{\log \frac{g(\boldsymbol{y})}{f(\boldsymbol{y},\boldsymbol{\xi})}\Big\}, \tag{2}$$

where the expectation is taken with respect to the true distribution.

Note that if the model for $\boldsymbol{y}$ is correctly specified, then the information criterion attains its unique minimum at $\boldsymbol{\xi}^* = \boldsymbol{\xi}_0$. In such a case, $\hat{\boldsymbol{\xi}}_n$ is a consistent estimator for $\boldsymbol{\xi}_0$ and the inverse of the Fisher Information-Matrix can be used to obtain standard errors for $\hat{\boldsymbol{\xi}}_n$. However, this matrix does not yield valid results when the model is misspecified, thence appropriate standard errors can be obtained by replacing the asymptotic covariance matrix by the so-called sandwich estimator. To this end, we introduce the following additional notation

$$\begin{aligned}
\mathbf{A}(\boldsymbol{\xi}) &= \mathsf{E}\Big\{\frac{\partial^2 \log f(\boldsymbol{y}_t,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k\partial\boldsymbol{\xi}_\ell}\Big\}, \\
\mathbf{B}(\boldsymbol{\xi}) &= \mathsf{E}\Big\{\frac{\partial \log f(\boldsymbol{y}_t,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k}\cdot\frac{\partial \log f(\boldsymbol{y}_t,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_\ell}\Big\}, \\
\mathbf{A}_n(\boldsymbol{\xi}) &= \Big\{\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \log f(\boldsymbol{y}_t,\boldsymbol{\xi})}{\partial\boldsymbol{\xi}_k\partial\boldsymbol{\xi}_\ell}\Big\},
\end{aligned}$$

$$\mathbf{B}_n(\boldsymbol{\xi}) = \Big\{\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \log f(\boldsymbol{y}_t,\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_k}\cdot\frac{\partial \log f(\boldsymbol{y}_t,\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_\ell}\Big\},$$

for $k, \ell = 1, \ldots, p$, where p refers to the number of parameters in the model. Further, let us define $\mathbf{V}(\boldsymbol{\xi}) = \mathbf{A}^{-1}(\boldsymbol{\xi})\mathbf{B}(\boldsymbol{\xi})\mathbf{A}^{-1}(\boldsymbol{\xi})$. White (1982) proved that, in general, $\hat{\boldsymbol{\xi}}_n \sim \mathrm{N}(\boldsymbol{\xi}^*, \mathbf{V}(\boldsymbol{\xi}^*))$. Additionally, he showed that, under a correctly specified model, $\mathbf{B}(\boldsymbol{\xi}_0)+\mathbf{A}(\boldsymbol{\xi}_0) = \mathbf{0}$, and therefore,

$$\mathbf{V}(\boldsymbol{\xi}_0) = -\mathbf{A}^{-1}(\boldsymbol{\xi}_0), \tag{3}$$

i.e., we then recover the inverse of the Fisher Information-Matrix.

The Information-Matrix Test (IMT), introduced by White (1982), is based on the elements of $\mathbf{B}(\boldsymbol{\xi}) + \mathbf{A}(\boldsymbol{\xi})$, which can be estimated using $\mathbf{B}_n(\hat{\boldsymbol{\xi}}_n) + \mathbf{A}_n(\hat{\boldsymbol{\xi}}_n)$. As discussed by this author, it may be prohibitive to base the test on all elements of this matrix. So, for simplicity reasons, we will focus on the diagonal elements.

Following the notation of White (1982), let $\boldsymbol{d}(\boldsymbol{y}, \boldsymbol{\xi})$ represent the $p \times 1$ vector with elements

$$d_k(\boldsymbol{y},\boldsymbol{\xi}) = \Big\{\frac{\partial \log f(\boldsymbol{y},\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_k}\Big\}^2 + \frac{\partial^2 \log f(\boldsymbol{y}_t,\boldsymbol{\xi})}{(\partial \boldsymbol{\xi}_k)^2},$$

$k = 1, \ldots, p$. Then, $\boldsymbol{d}_n(\hat{\boldsymbol{\xi}}_n) = n^{-1}\sum_{t=1}^{n}\boldsymbol{d}(\boldsymbol{y}_t, \hat{\boldsymbol{\xi}}_n)$ represents the $p \times 1$ vector containing the diagonal elements of $\mathbf{B}_n(\hat{\boldsymbol{\xi}}_n) + \mathbf{A}_n(\hat{\boldsymbol{\xi}}_n)$. Further, let

$$\nabla \boldsymbol{d}_n(\boldsymbol{\xi}_k) = \Big\{\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{d}(\boldsymbol{y}_t,\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_k}\Big\}, \tag{4}$$

$(k = 1, \ldots, p)$ and define

$$\begin{aligned}\mathbf{C}_n(\hat{\boldsymbol{\xi}}_n) = {} & \frac{1}{n}\sum_{t=1}^{n}[\boldsymbol{d}(\boldsymbol{y}_t,\hat{\boldsymbol{\xi}}_n) - \nabla \boldsymbol{d}_n(\hat{\boldsymbol{\xi}}_n)\mathbf{A}_n^{-1}(\hat{\boldsymbol{\xi}}_n)\nabla \log f(\boldsymbol{y}_t,\hat{\boldsymbol{\xi}}_n)] \\ & \times[\boldsymbol{d}(\boldsymbol{y}_t,\hat{\boldsymbol{\xi}}_n) - \nabla \boldsymbol{d}_n(\hat{\boldsymbol{\xi}}_n)\mathbf{A}_n^{-1}(\hat{\boldsymbol{\xi}}_n)\nabla \log f(\boldsymbol{y}_t,\hat{\boldsymbol{\xi}}_n)]^T.\end{aligned}$$

Using these elements, the Information-Matrix Test statistic is defined as

$$\Im_n = n\boldsymbol{d}_n^T(\hat{\boldsymbol{\xi}}_n)\mathbf{C}_n(\hat{\boldsymbol{\xi}}_n)^{-1}\boldsymbol{d}_n(\hat{\boldsymbol{\xi}}_n)$$

and follows, under a correctly specified model, a χ^2 distribution with p degrees of freedom (White, 1982).

As can be seen from (4), the calculation of $\Im_n$ requires the third order partial derivatives of the log-likelihood. Since in GLMM the likelihood is not available in closed form, these derivatives cannot be obtained analytically and hence one has to resort to numerical approximations, which can be burdensome and less than straightforward to carry out using conventional statistical packages, even when GLMM models exist (Molenberghs and Verbeke, 2005). This further underscores the need for sensible diagnostic tools that perform equally well as the available misspecification test, but at the same time avoid higher order partial derivatives. In the following section, such an alternative is proposed.

4. THE SANDWICH-ESTIMATOR TEST

Let us recall that, under a correctly specified model, (3) holds and hence deviations from the model assumptions are expected to distort this equality. Our proposal focuses on the difference between $\mathbf{V}(\boldsymbol{\xi})$ and $-\mathbf{A}^{-1}(\boldsymbol{\xi})$ as a potential indicator of misspecification. Similarly as before, we will consider the diagonal elements only, i.e., $\text{diag}[\mathbf{V}(\boldsymbol{\xi}) + \mathbf{A}^{-1}(\boldsymbol{\xi})]$.

First, let us define $\mathbf{V}_n(\boldsymbol{\xi}) = \mathbf{A}^{-1}(\boldsymbol{\xi})\mathbf{B}_n(\boldsymbol{\xi})\mathbf{A}^{-1}(\boldsymbol{\xi})$, and let $\boldsymbol{v}_n(\boldsymbol{\xi}) = \text{diag}[\mathbf{V}_n(\boldsymbol{\xi}) + \mathbf{A}^{-1}(\boldsymbol{\xi})]$. Note that $\boldsymbol{v}_n(\boldsymbol{\xi})$ can also be written as $\boldsymbol{\Delta}\text{vec}[\mathbf{V}_n(\boldsymbol{\xi}) + \mathbf{A}^{-1}(\boldsymbol{\xi})]$, where $\boldsymbol{\Delta}$ is the $p \times p^2$ matrix specified as

$$\boldsymbol{\Delta} = \begin{cases} 1 & \text{for } k = 1, \ldots, p \text{ and } \ell = (k-1)p + i, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

An estimator for the covariance $\mathbf{C}_v(\boldsymbol{\xi})$ of $\boldsymbol{v}_n(\boldsymbol{\xi})$ is therefore given by

$$\mathbf{C}_v(\boldsymbol{\xi}) = \text{cov}\{\boldsymbol{\Delta}\text{vec}[\mathbf{V}_n(\boldsymbol{\xi}) + \mathbf{A}^{-1}(\boldsymbol{\xi})]\} = \boldsymbol{\Delta}\text{cov}\{\text{vec}[\mathbf{V}_n(\boldsymbol{\xi})]\}\boldsymbol{\Delta}^T. \tag{6}$$

Using some properties of the Kronecker product and the vec operator, it is possible to show that

$$\text{vec}[\mathbf{V}_n(\boldsymbol{\xi})] = [\mathbf{A}^{-1}(\boldsymbol{\xi}) \otimes \mathbf{A}^{-1}(\boldsymbol{\xi})]\text{vec}[\mathbf{B}_n(\boldsymbol{\xi})]$$

and therefore

$$\text{cov}\{\text{vec}[\mathbf{V}_n(\boldsymbol{\xi})]\} = [\mathbf{A}^{-1}(\boldsymbol{\xi}) \otimes \mathbf{A}^{-1}(\boldsymbol{\xi})]\text{cov}\{\text{vec}[\mathbf{B}_n(\boldsymbol{\xi})]\}[\mathbf{A}^{-1}(\boldsymbol{\xi}) \otimes \mathbf{A}^{-1}(\boldsymbol{\xi})]^T.$$

This implies that the problem of finding an estimator for the covariance of $\boldsymbol{v}_n(\boldsymbol{\xi})$ reduces to finding an estimator for the covariance of vec$[\mathbf{B}_n(\boldsymbol{\xi})]$. Let us now write

$$\boldsymbol{p}_t(\boldsymbol{\xi}) = \text{vec}\left(\frac{\partial \log f(\boldsymbol{y}_t, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}_k} \cdot \frac{\partial \log f(\boldsymbol{y}_t, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}_\ell}\right),$$

and let $\mathbf{P}(\boldsymbol{\xi})$ and $\mathbf{T}(\boldsymbol{\xi})$ represent the mean and the covariance of $\boldsymbol{p}_t(\boldsymbol{\xi})$, respectively. An unbiased estimator for $\mathbf{P}(\boldsymbol{\xi})$ is given by $\bar{\boldsymbol{p}}(\boldsymbol{\xi}) = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{p}_t(\boldsymbol{\xi})$ and $\mathbf{T}(\boldsymbol{\xi})$ can be estimated using the empirical covariance estimator

$$\hat{\mathbf{T}}(\boldsymbol{\xi}) = \frac{1}{n-1}\sum_{t=1}^{n}[\boldsymbol{p}_t(\boldsymbol{\xi}) - \bar{\boldsymbol{p}}(\boldsymbol{\xi})][\boldsymbol{p}_t(\boldsymbol{\xi}) - \bar{\boldsymbol{p}}(\boldsymbol{\xi})]^T.$$

Note further that vec$[\mathbf{B}_n(\boldsymbol{\xi})] = \bar{\boldsymbol{p}}(\boldsymbol{\xi})$ and therefore cov$\{$vec$[\mathbf{B}_n(\boldsymbol{\xi})]\} = n^{-1}\hat{\mathbf{T}}(\boldsymbol{\xi})$. From the previous expressions, we obtain

$$\text{cov}\{\text{vec}[\mathbf{V}_n(\boldsymbol{\xi})]\} = \frac{1}{n}[\mathbf{A}^{-1}(\boldsymbol{\xi}) \otimes \mathbf{A}^{-1}(\boldsymbol{\xi})]\hat{\mathbf{T}}(\boldsymbol{\xi})[\mathbf{A}^{-1}(\boldsymbol{\xi}) \otimes \mathbf{A}^{-1}(\boldsymbol{\xi})].$$

Finally, the following theorem defines the test statistic and establishes its distribution under the null hypothesis.

Theorem 1 *Under general regularity conditions when the model is correctly specified*

$$\boldsymbol{v}_n(\boldsymbol{\xi}_0) \sim N_p(\mathbf{0}, \mathbf{C}_v(\boldsymbol{\xi}_0)),$$

and therefore

$$\delta_n = \boldsymbol{v}_n^T(\boldsymbol{\xi}_0)[C(\boldsymbol{\xi}_0)]^{-1}\boldsymbol{v}_n(\boldsymbol{\xi}_0) \sim \chi_p^2. \tag{7}$$

When Theorem 1 is applied in a practical situation, $\boldsymbol{\xi}_0$ and $\mathbf{A}^{-1}(\boldsymbol{\xi}_0)$ in (7) can be substituted by their consistent estimators under the null, given by $\hat{\boldsymbol{\xi}}_n$ and $\mathbf{A}^{-1}(\hat{\boldsymbol{\xi}}_n)$ respectively. This test can be implemented in a comparatively easy way using standard statistical software, like the procedures NLMIXED and IML in SAS. A SAS macro that performs all of the necessary calculations is available on the authors' web pages.

Clearly, substituting $\boldsymbol{\xi}_0$ and $\mathbf{A}^{-1}(\boldsymbol{\xi}_0)$ by consistent estimators in (7) introduces a certain amount of extra variability not taken into account by the test; the price to pay for the gain in simplicity. We will assess these and other issues by way of a simulation study, to be reported next.

5. SIMULATION STUDY

To explore the impact of the aforementioned substitutions, as well as the small sample size behavior of the test, we designed a simulation study to compare the performance of both tests, IMT and SET, within the LMM framework with misspecified random-effects distribution. This is feasible since for LMM a closed-form likelihood is available and all the derivatives involved in the IMT can be calculated analytically. Additionally, a second simulation study was designed to explore empirically the power and type I error rate of the new test in GLMM, where IMT is difficult to apply, under random-effects misspecification.

5.1 Linear Mixed Models

The data were generated using a linear random-intercept model

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 z_i + \beta_2 t_j + \varepsilon_{ij}, \tag{8}$$

including an intercept, a binary covariate z_i taking values 0 and 1, a within-cluster covariate t_j taking values 0, 1, 2, 4, 6, and 8, measurement error ε_{ij} sampled from a standard normal distribution $N(0,1)$, and a random intercept b_{0i} sampled from 5 distinct mean-zero distributions, each with variances $\sigma_b^2 = 4$ and 32: the normal, power function, and lognormal distributions, as well as a discrete distribution with equal probability at two support points, and an asymmetric mixture of two normal densities. Note that $\sigma_b^2 = 32$ is of the same order of magnitude as the estimate obtained from 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 $\beta_2 = 1$, also in accordance with the values estimated from the case study. Six different sample sizes were considered, namely 50, 100, 200, 350, 500, and 1000. For each setting, 500 data sets were generated and Model (8) was fitted to these data under the assumption of normally distributed random effects. We then determined the proportion of cases in which a result significant at the

5% significance level was detected, using both $\Im_n$ and δ_n. When the random effects are generated from a normal distribution, this proportion corresponds to the type I error; in the other settings, it represents the power of the tests. The results of these simulations are shown in the first panel of Table 2.

Clearly, the IMT and the newly proposed test have a very similar performance in all settings considered. This suggests that using $\hat{\boldsymbol{\xi}}_n$ and $\mathbf{A}^{-1}(\hat{\boldsymbol{\xi}}_n)$ in (7) does not have a big impact on the asymptotic behavior of the SET, at least relative to the IMT, which does take into account all sources of variability. Combined with its computational advantage, this is a strong point for the SET.

In general, both tests show a good power and type I error rate for reasonable sample sizes ($n = 350$). The only exception to the previous behavior is observed when the true distribution of the random effects is an asymmetric mixture with small variance. Nevertheless, it should be noted that for small variances this asymmetric mixture can be reasonably well approximated by a normal density. As a consequence, both tests need a relatively large sample size to detect the heterogeneity.

Finally, we should like to point out that, even though Verbeke and Lesaffre (1997) showed that maximum likelihood estimators in LMM are still consistent and asymptotically normal when the random-effects distribution is misspecified, this asymptotic convergency can be either fast or slow, depending on the severity of the departure from normality. It would therefore be practically relevant to use diagnostic tools like the IMT or SET in this scenario as well. Let us now turn to the GLMM case.

5.2 Generalized Linear Mixed Models

In these simulations, binary responses were generated using the logistic random-intercept model given by (1). The same random-effects distributions and parameter values as in the previous simulations were used. The results are reported on the right hand side of Table 2.

Here again, a good power and type I error rate is observed with a reasonable sample size ($n = 350$). Clearly the test performs better when the variance of the random effects is large. This is however, a desirable behavior given the results obtained by Litière *et al.* (2006) and presented in Section 1.

Like for LMM, the asymmetric mixture appears to be a misspecification difficult to detect. Not surprisingly, in GLMM the power of SET is even lower in this case than in the corresponding LMM situation, because a binary response conveys much less information than a continuous one. Overall though, the SET can be considered a useful tool to evaluate the appropriateness of the model.

6. ANALYSIS OF CASE STUDY

In this section we will apply the SET to assess the suitability of Model (1) with normal random effects for the analysis of the case study. It follows that $\delta_n = 0.445$ and compared to a χ^2 distribution with 4 degrees of freedom, this leads to $p = 0.979$. This implies that the data at hand do not give evidence against the assumption of normally distributed random effects. However, it should also be noted that, according to our simulations, the sample size available might not be enough to achieve a good level of power.

7. DISCUSSION

A commonly encountered perception among data analysts is that the choice of the random-effects distribution in generalized linear mixed models is not crucial to the quality of the inferences drawn. However, recent research is showing that the maximum likelihood estimators and the inferential procedures can be affected. This emphasizes the need for diagnostic tools to detect model misspecification. In this manuscript, we proposed the so-called Sandwich-Estimator Test to detect random-effects misspecification in generalized linear mixed models. We compared, via simulations, the performance of the new test with the performance of the relatively well-known Information-Matrix Test of White (1982), in the context of linear mixed models. We found that both tests performed very similar. However, a strong advantage of our proposal is that it does not require third order partial derivatives of the log-likelihood function, making it easy to implement in models where the likelihood is not available in closed form, such as, importantly, in the generalized linear mixed model. Encouragingly, simulations have shown that the SET performs very well in situations where the random-effects misspecification can have a serious impact on the maximum likelihood estimators.

Even though the test was initially developed to detect random-effects misspecification, we should like to point out that this test is suitable to detect also other types of model misspecification like, for example, the choice of the link function, the mean structure etc. Further research is required to establish the SET's performance to detect other types of model misspecification.

References

Agresti, A., Caffo, B., and Ohman-Strickland, P. (2004), "Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies," *Computational Statistics and Data Analysis*, 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.

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.

Litière, S., Alonso, A., and Molenberghs, G. (2007), "Type I and type II error random-effects misspecification in generalized linear mixed models," *Biometrics*, 00, 000–000.

Litière, S., Alonso, A., Molenberghs, G., and Geys, H. (2006), "The impact of a misspecified random-effects distribution on maximum likelihood estimation in generalized linear mixed models," *Submitted for publication*.

Molenberghs, G. and Verbeke, G. (2005), *Models for Discrete Longitudinal Data*, New York: Springer.

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.

Verbeke, G. and Lesaffre, E. (1997), "The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data," *Computational Statistics and Data Analysis*, 53, 541–556.

White, H. (1982), "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1-25.

Table 1: *Parameter estimates (standard errors) and p-values for the fixed effects in the logistic random-intercept Model (1), fitted to the schizophrenia data.*

Effect	Parameter	Estimate (s.e.)	p-value
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

Table 2: *Power and type I error for detecting a misspecified random-effects distribution, using $\Im_n$ and δ_n in the setting of linear mixed models, and using δ_n in generalized linear mixed models.*

		LMM				GLMM	
		$\sigma_b^2 = 4$		$\sigma_b^2 = 32$		$\sigma_b^2 = 4$	$\sigma_b^2 = 32$
Distribution	n	$\Im_n$	δ_n	$\Im_n$	δ_n	δ_n	δ_n
Normal	50	0.266	0.298	0.232	0.284	0.278	0.234
	100	0.208	0.204	0.190	0.172	0.192	0.122
	200	0.106	0.088	0.116	0.090	0.098	0.072
	350	0.084	0.058	0.096	0.060	0.046	0.038
	500	0.048	0.034	0.078	0.064	0.044	0.040
	1000	0.042	0.026	0.054	0.040	0.026	0.016
Power function	50	0.372	0.458	0.368	0.482	0.242	0.264
	100	0.432	0.490	0.460	0.516	0.172	0.288
	200	0.640	0.636	0.676	0.686	0.092	0.406
	350	0.778	0.780	0.848	0.848	0.086	0.650
	500	0.876	0.864	0.888	0.886	0.096	0.812
	1000	0.958	0.952	0.974	0.972	0.164	0.998
Discrete	50	1.000	1.000	1.000	1.000	0.304	0.710
	100	1.000	1.000	1.000	1.000	0.306	0.830
	200	1.000	1.000	1.000	1.000	0.430	0.948
	350	1.000	1.000	1.000	1.000	0.688	0.996
	500	1.000	1.000	1.000	1.000	0.848	1.000
	1000	1.000	1.000	1.000	1.000	0.986	1.000
Asymmetric mixture	50	0.136	0.228	0.562	0.660	0.230	0.242
	100	0.156	0.182	0.562	0.592	0.146	0.172