



Detecting random-effects model misspecification via coarsened data

Xianzheng Huang*

Department of Statistics, University of South Carolina, Columbia, SC 29208, USA

ARTICLE INFO

Article history:

Received 29 June 2009

Received in revised form 10 June 2010

Accepted 11 June 2010

Available online 20 June 2010

Keywords:

Generalized linear mixed models

Kullback–Leibler divergence

Nonlinear mixed models

ABSTRACT

Mixed effects models provide a suitable framework for statistical inference in a wide range of applications. The validity of likelihood inference for this class of models usually depends on the assumptions on random effects. We develop diagnostic tools for detecting random-effects model misspecification in a rich class of mixed effects models. These methods are illustrated via simulation and application to soybean growth data.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Mixed effects models are widely used in statistical applications in biology, agriculture, sociology, and environmental science, where clustered data are often collected. Under the framework of mixed effects models, it is straightforward to make prediction at the inter-cluster level as well as the intra-cluster level. [Davidian and Giltinan \(2003\)](#) and [McCulloch et al. \(2008\)](#) provide a comprehensive survey of mixed effects models. A major concern in this realm lies in random-effects assumptions. [Hartford and Davidian \(2000\)](#), [Heagerty and Kurland \(2001\)](#), [Agresti et al. \(2004\)](#) and [Litière et al. \(2007, 2008\)](#) showed various adverse effects of erroneous random-effects assumptions on inference.

Semiparametric and nonparametric methods have been developed to relax the parametric assumptions on random effects. [Davidian and Gallant \(1993\)](#) used a flexible, smooth density to characterize the distribution of random effects. [Fattinger et al. \(1995\)](#) used spline functions to transform the random effects, initially assumed to be normal, so that the resulting random effects can follow an arbitrary distribution dictated by data. [Lai and Shih \(2003\)](#) took a nonparametric approach to estimate the distribution of random effects. The price one pays to avoid suspicious parametric assumptions by using these methods is often heavy computation. Some of these methods only provide discrete estimation for the random-effects distribution. This can be unsatisfactory when the distributional characteristics of random effects are of interest. Because a parsimonious parametric model can be appealing for its simplicity and potential gain in efficiency, techniques to check parametric assumptions on random effects in mixed effects models are valuable.

There are three main lines of development thus far for assessing the validity of model specification. The first line of work is based on the information matrix equivalence under a correct model ([White, 1981, 1982](#)). Focusing on the variance–covariance structure for the random effects in nonlinear mixed models (NLMM), [Vonesh et al. \(1996\)](#) compared two variance–covariance estimators that are equivalent under the correct model. [Litière \(2007\)](#) developed a series of tests motivated by the information matrix equivalence to diagnose random-effects misspecification in linear mixed models (LMM) and generalized linear mixed models (GLMM). The second line of development aims at obtaining an empirical distribution of the random effects. For example, [Lange and Ryan \(1989\)](#) estimated the cumulative distribution of random effects using the empirical Bayes estimates of individual random effects. [Ritz \(2004\)](#) extended their results and derived a

* Tel.: +1 803 777 8772; fax: +1 803 777 4048.

E-mail address: huang@stat.sc.edu.

weighted empirical process for the random effects. Waagepetersen (2006) simulated random effects from the conditional distribution of random effects given the observed data in GLMM. The third line of research has some similarity with the first line but is relatively underdeveloped. Instead of comparing two information matrices, one compares two estimators for the fixed effects, with one robust to model misspecification and the other sensitive to it. This idea was once prevalent in the econometrics community (Hausman, 1978) and was recently used by Tchetgen and Coull (2006) to construct a test for a specific form of GLMM.

These existing diagnostic methods are either too general to provide guidance for correction when misspecification is detected or too specific to allow immediate extension to complex models or to detect departures other than normality assumption. Admittedly, as shown by Verbeke and Molenberghs (submitted for publication), without assuming other component models in a hierarchical mixed effects model correctly specified, random-effects assumptions are unverifiable. With multiple uncertain component models, one can test the sensitivity of inference to multiple model assumptions but may not be able to single out the assumptions on random effects. If one has more confidence in the other component models than in the random effects, then it is possible to verify random-effects assumptions by studying the robustness of inference under these assumptions. This is the underlying philosophy of the methods discussed in this article. As a starting point, Huang (2009) provided empirical evidence that MLEs in a GLMM for binary response are not robust to data grouping in the presence of random-effects model misspecification. Motivated by this finding, she constructed diagnostic tests based on the discrepancy between the observed-data MLEs and the MLEs from the induced grouped data. Compared to the aforementioned existing methods, her tests can be more informative because the outcomes of the tests can depend on the source of model misspecification besides its existence.

In this article, we shed further light on the idea motivating Huang's method, which is the content of Section 2. In Section 3, this idea is extended to a richer class of mixed effects models where the response is not binary. New test statistics that are computationally more efficient than those in Huang (2009) are defined in Section 4, where simulation studies are also presented. In Section 5, the diagnostic techniques are applied to a real data example. Concluding remarks and future research topics are given in Section 6.

2. Diagnostic method for GLMM

2.1. Test statistics

Denote by Ω the $p \times 1$ vector of unknown parameters in the model, by $\tilde{\Omega}$ and $\tilde{\Omega}_c$ the limiting MLE based on the observed data and that based on the grouped data, respectively, where the limit is taken as the number of clusters, m , tends to infinity and the size of the cluster is bounded. Under a correct model, it is expected that $\tilde{\Omega} = \tilde{\Omega}_c$, and erroneous model assumptions can result in $\tilde{\Omega} \neq \tilde{\Omega}_c$. Following this claim, Huang (2009) considered testing $H_0 : \tilde{\Omega} = \tilde{\Omega}_c$, and proposed test statistics that compare the finite-sample counterparts of $\tilde{\Omega}$ and $\tilde{\Omega}_c$, denoted by $\hat{\Omega}$ and $\hat{\Omega}_c$. The follow test statistics implement elementwise comparison,

$$\mathbf{t}_1 = (\hat{\Omega} - \hat{\Omega}_c) \# \text{vecdiag}(\hat{\mathbf{V}}_1^{-1}), \quad (1)$$

where $\hat{\mathbf{V}}_1$ is an estimator for the variance-covariance matrix of $\hat{\Omega} - \hat{\Omega}_c$, $\text{vecdiag}(\hat{\mathbf{V}}_1^{-1})$ denotes the column vector consisting of the diagonal elements of $\hat{\mathbf{V}}_1^{-1}$, and “#” is the elementwise multiplication operator. Detailed derivation of $\hat{\mathbf{V}}_1$ is given in Huang (2009), where it is shown that, under H_0 , an element in \mathbf{t}_1 associated with a parameter γ , $t_1(\gamma)$, follows a Student's t distribution with $m - p$ degrees of freedom. Another test statistic compares the entire vector of MLEs as follows,

$$F_1 = \frac{m - p}{p(m - 1)} (\hat{\Omega} - \hat{\Omega}_c)^T \hat{\mathbf{V}}_1^{-1} (\hat{\Omega} - \hat{\Omega}_c), \quad (2)$$

which follows an $F(p, m - p)$ distribution under H_0 . A significantly large value of F_1 casts doubt about the veracity of H_0 . Individual values of $t_1(\gamma)$ assess the sensitivity of individual parameter estimate to data grouping, which can relate to how a model is misspecified. These tests do not directly test the validity of random-effects assumptions. Rather, they are designed to test if the inference is robust to data grouping under these model assumptions, and being robust is no guarantee for correct model assumptions, but is merely some reassurance that the inference, even if it is inconsistent, may not deteriorate (in terms of consistency but not efficiency) when grouped data are used. In the next subsection, we use a concrete example to elaborate the fundamental idea behind these tests and, by so doing, give one some confidence in extending this idea to other mixed effects models.

2.2. Limiting maximum likelihood estimators and coarsened data

Grouped data is a special case of coarsened data (Heitjan and Rubin, 1991; Tsiatis, 2006). Other examples of coarsened data common in practice include censored data, incomplete data due to missingness, rounded data, etc. Before focusing on grouped data, it is instructive to first introduce the notion of coarsened data generically. Let $(\Delta_i, \mathbf{Y}_i^*)$ be the i th datum in the coarsened data, where Δ_i is the coarsening variable (if needed), and $\mathbf{Y}_i^* = C_{\Delta_i}(\mathbf{Y}_i)$ is the coarsened response, $C_{\Delta_i}(\mathbf{Y}_i)$ is

a many-to-one coarsening function that maps \mathbf{Y}_i to \mathbf{Y}_i^* . Designing a coarsening mechanism includes specifying a probability model for Δ_i and defining a coarsening function $C_{\Delta_i}(\cdot)$. Likelihood inference based on coarsened data can depend on both aspects of a coarsening mechanism. Moreover, the effect of random-effects model misspecification on inference based on coarsened data can also depend on these two aspects. In other words, likelihood inference based on coarsened data can be affected by the interaction of model misspecification and coarsening mechanism. In what follows, an example is used to demonstrate explicitly that $\tilde{\Omega}_c$ resulting from grouped data depends on the group size and the source of misspecification.

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ denote the vector of binary responses observed from cluster i , for $i = 1, \dots, m$. Suppose that one assumes a logistic model as follows,

$$P(Y_{ij} = 1 | \mathbf{X}_{ij}, b_{i0}) = \{1 + \exp(-\beta_0 - b_{i0} - \beta_1 X_{ij,1} - \beta_2 X_{ij,2} - \beta_3 X_{ij,1} X_{ij,2})\}^{-1}, \tag{3}$$

where $\mathbf{X}_{ij} = (X_{ij,1}, X_{ij,2})^T$, $X_{ij,1} = x_i$ ($=0$ or 1) is a between-cluster covariate, $X_{ij,2} = (j - 1)/(n - 1)$ is a within-cluster covariate, for $j = 1, \dots, n_i$, b_{i0} is a random intercept, $n_i = n = 12$ for $i = 1, \dots, m$, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ is the vector of fixed effects. Moreover, one assumes that b_{i0} , $i = 1, \dots, m$, are independent and identically distributed according to $N(0, \sigma_0^2)$. Under these assumptions, the observed-data density associated with cluster i is given by

$$f_{\mathbf{Y}_i}(\mathbf{Y}_i | \mathbf{X}_i; \boldsymbol{\Omega}) = \int f_{b_{i0}}(b_{i0}; \sigma_0) \prod_{j=1}^{n_i} P(Y_{ij} = 1 | \mathbf{X}_{ij}, b_{i0})^{Y_{ij}} P(Y_{ij} = 0 | \mathbf{X}_{ij}, b_{i0})^{1-Y_{ij}} db_{i0}, \tag{4}$$

where $f_{b_{i0}}(b_{i0}; \sigma_0) = \sigma_0^{-1} \phi(b_{i0}/\sigma_0)$, $\phi(\cdot)$ is the standard normal density function, and $\boldsymbol{\Omega} = (\boldsymbol{\beta}^T, \sigma_0)^T$.

Given the above assumed (possibly misspecified) model, we next design three true models that deviate from the assumed model in different ways to represent several commonly encountered misspecification in practice. Then we will demonstrate the dependence of $\tilde{\Omega}_c$ on data grouping under each type of misspecification. Denote by $g_{\mathbf{Y}_i}(\mathbf{Y}_i | \mathbf{X}_i; \boldsymbol{\Omega}_0)$ the true observed-data density for cluster i , where $\boldsymbol{\Omega}_0$ is the vector of true parameters, among which $\boldsymbol{\beta} = (-2, 1, 0.5, -0.25)^T$.

- (M1) $g_{\mathbf{Y}_i}(\mathbf{Y}_i | \mathbf{X}_i; \boldsymbol{\Omega}_0)$ is identical to (4) except that $b_{i0} = \sigma_0(a_i - \lambda)/\sqrt{\lambda}$, where $a_i \sim \text{gamma}(\lambda, 1)$, $\sigma_0 = 3$, and $\lambda = 1$.
- (M2) $g_{\mathbf{Y}_i}(\mathbf{Y}_i | \mathbf{X}_i; \boldsymbol{\Omega}_0)$ is identical to (4) except that $[b_{i0} | x_i = 0] \sim N(0, \sigma_{00}^2)$ and $[b_{i0} | x_i = 1] \sim N(0, \sigma_{01}^2)$, where $\sigma_{00} = 3$ and $\sigma_{01} = 0.5$.
- (M3) The true logistic model is given by

$$P(Y_{ij} = 1 | \mathbf{X}_{ij}, b_{i0}, b_{i1}) = \{1 + \exp[-\beta_0 - b_{i0} - \beta_1 X_{ij,1} - (\beta_2 + b_{i1}) X_{ij,2} - \beta_3 X_{ij,1} X_{ij,2}]\}^{-1},$$

where $b_{i0} \sim N(0, \sigma_0^2)$ is independent of $b_{i1} \sim N(0, \sigma_1^2)$, $\sigma_0 = 0.5$, and $\sigma_1 = 2$.

Exploiting the result in [Rotnitzky and Wypij \(1994\)](#), one has

$$\tilde{\boldsymbol{\Omega}} = \arg \max_{\boldsymbol{\Omega}} \sum_{l=1}^{2^n} \{ \pi^{(1)}(\mathbf{Y}_l) \log f_{\mathbf{Y}_i}(\mathbf{Y}_l | \mathbf{X}^{(1)}; \boldsymbol{\Omega}) + \pi^{(2)}(\mathbf{Y}_l) \log f_{\mathbf{Y}_i}(\mathbf{Y}_l | \mathbf{X}^{(2)}; \boldsymbol{\Omega}) \}, \tag{5}$$

where $\{\mathbf{Y}_l, l = 1, \dots, 2^n\}$ forms the sample space of the $n \times 1$ binary response \mathbf{Y} , $\mathbf{X}^{(1)} = [\mathbf{1} \ \mathbf{0} \ \mathbf{S} \ \mathbf{0}]$ and $\mathbf{X}^{(2)} = [\mathbf{1} \ \mathbf{1} \ \mathbf{S} \ \mathbf{S}]$ are the two distinct fixed-effects design matrices, $\mathbf{1}$ denotes the $n \times 1$ vector of ones, $\mathbf{0}$ denotes the $n \times 1$ vector of zeros, and $\mathbf{S} = (0, 1/(n - 1), 2/(n - 1), \dots, 1)^T$. Lastly, $\pi^{(k)}(\mathbf{Y}_l) = g_{\mathbf{Y}_i}(\mathbf{Y}_l | \mathbf{X}^{(k)}; \boldsymbol{\Omega}_0)$, for $k = 1, 2$, are the true probabilities of observing \mathbf{Y}_l when the fixed-effects design matrix is $\mathbf{X}^{(k)}$, for $l = 1, \dots, 2^n$. Computation details to obtain $\tilde{\boldsymbol{\Omega}}$ are described in [Huang \(2009\)](#) and omitted here.

Based on the observed data, we generate grouped binary responses by first forming G_i groups within cluster i , then defining $\mathbf{Y}_i^* = (Y_{i1}^*, \dots, Y_{iG_i}^*)^T$, where Y_{ig}^* is equal to one if at least one of the Y_{ij} 's in group g is one, and zero otherwise, for $g = 1, \dots, G_i$, $i = 1, \dots, m$. For simplicity, we set $G_i = G$ for $i = 1, \dots, m$. The density of the grouped data, $f_{\mathbf{Y}_i^*}(\mathbf{Y}_i^* | \mathbf{X}_i; \boldsymbol{\Omega})$, can be straightforwardly derived based on $f_{\mathbf{Y}_i}(\mathbf{Y}_i | \mathbf{X}_i; \boldsymbol{\Omega})$ (see Eq. (3) in [Huang \(2009\)](#)). Then $\tilde{\boldsymbol{\Omega}}_c$ can be similarly defined by (5), with $f_{\mathbf{Y}_i}(\mathbf{Y}_i | \mathbf{X}^{(k)}; \boldsymbol{\Omega})$ replaced by $f_{\mathbf{Y}_i^*}(\mathbf{Y}_i^* | \mathbf{X}^{(k)}; \boldsymbol{\Omega})$, the true probabilities $\pi^{(k)}(\mathbf{Y}_i^*) = g_{\mathbf{Y}_i^*}(\mathbf{Y}_i^* | \mathbf{X}^{(k)}; \boldsymbol{\Omega}_0)$, and the new sample space being $\{\mathbf{Y}_l^*, l = 1, \dots, 2^G\}$. To check the accuracy of the calculation and to provide reference values for the limiting MLEs, we also compute $\tilde{\boldsymbol{\Omega}}$ and $\tilde{\boldsymbol{\Omega}}_c$ according to (5) when the true model coincides with the assumed model (4), as it is expected that $\tilde{\boldsymbol{\Omega}} = \tilde{\boldsymbol{\Omega}}_c = \boldsymbol{\Omega}_0$ in this case.

Under the case of correct modeling and three cases of model misspecification (M1)–(M3), we compute $\tilde{\boldsymbol{\Omega}}$ and $\tilde{\boldsymbol{\Omega}}_c$ with $G = 2, 3, 4, 6$, where $\boldsymbol{\Omega}$ can be viewed as a special case of $\boldsymbol{\Omega}_c$ with $G = n = 12$. The results are depicted in [Fig. 1](#). As expected, all MLEs are more or less biased in the presence of model misspecification, and the limiting MLEs under the correct modeling always retain at their true values. Note that, at a fixed level of G , the bias is invisible without knowing $\boldsymbol{\Omega}_0$; and the bias becomes visible as G varies (with $\boldsymbol{\Omega}_0$ still unknown) because then the bias also varies, as demonstrated in [Fig. 1](#). Since the value of G is part of the coarsening mechanism one designs, the dependence of the bias on G can be revealed by data at hand and so is the existence of bias. More importantly, [Fig. 1](#) suggests that which element in $\tilde{\boldsymbol{\Omega}}_c$ is most sensitive to the change in G depends on the source of model misspecification. In particular, in case (M1), where the only misspecification is on the distribution family of the random intercept, the dependence of $\tilde{\sigma}_{0c}$ on G stands out. In case (M2),

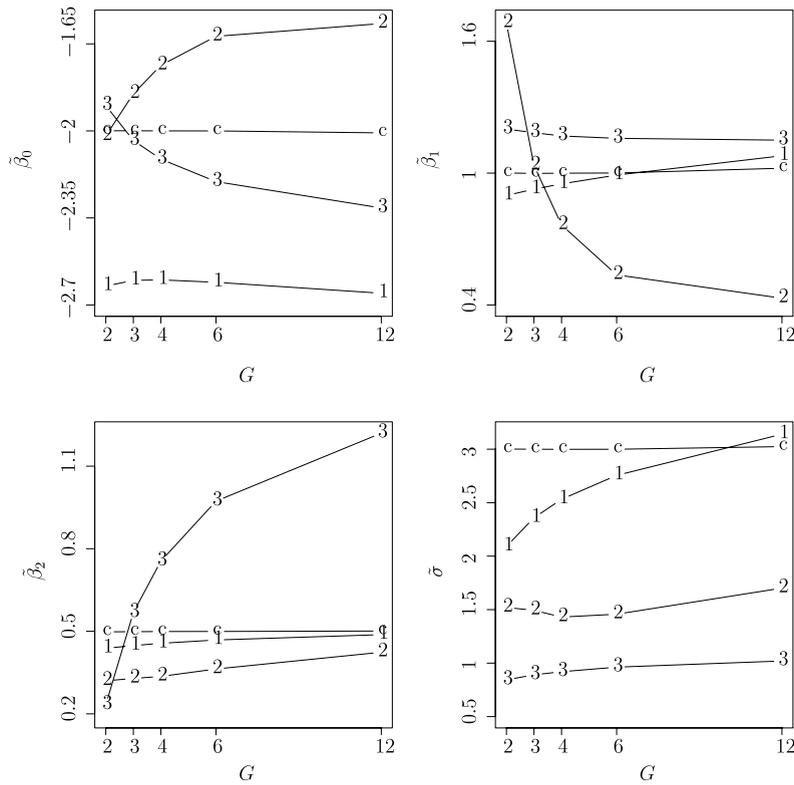


Fig. 1. Limiting MLEs of four parameters under the correct modeling (“c”) and under three cases of model misspecification (M1)–(M3) (“1”, “2”, and “3”) defined in Section 2.2. The group size of $G = 12$ refers to the observed data.

where the variance of the true random intercept depends on $X_{ij,1}$, but one ignores this heterogeneity in the assumed model, $\tilde{\beta}_{1c}$ changes most dramatically as G varies. In case (M3), where there is a random slope associated with $X_{ij,2}$ in the true model while one excludes it in the assumed model, $\tilde{\beta}_{2c}$ is the least robust to G .

In summary, the novel thrust of Huang’s method is to use coarsened data to reveal the dependence of the bias in MLEs on the interaction of data coarsening and model misspecification, and to further pinpoint the possible source of misspecification. This method is easy to implement because calculating the test statistics only involves routine maximum likelihood. The operation of grouping in this method forms a partition on the response vector space. Such partition differs from the partition on the covariate space used in the goodness-of-fit (GOF) test for logistic models proposed by Tsiatis (1980), the GOF tests for proportional hazards models discussed in Schoenfeld (1980), and the tests for GEE developed by Barnhart and Williamson (1998). Another type of partition used in constructing GOF tests is the one based on the predicted probabilities in logistic models, as in Hosmer and Lemeshow (1980) and Horton et al. (1999), the latter of which pointed out the equivalence between their partition and the partition on the covariate space. Besides the nature of the utilized partition, another distinct feature of our method is that we compare two sets of estimators for model parameters whereas the above existing GOF tests compare the response resulting from the grouping with the expected response such as expected cell counts.

Direct computation of $\tilde{\Omega}$ and $\tilde{\Omega}_c$ is infeasible for more complex data structure and models, such as NLMM. Hence explicit analyses of effects of model misspecification on $\tilde{\Omega}_c$ as above cannot be easily generalized. But it is not unreasonable to expect that, in general, the limiting MLE can be sensitive to data coarsening in the presence model misspecification. This motivates the following extension to NLMM for non-binary response.

3. Coarsened data in NLMM

3.1. Model specification

Now consider an NLMM with the intra-cluster model given by $Y_{ij} = f(\mathbf{X}_{ij}, \boldsymbol{\beta}_i) + \epsilon_{ij}$, $i = 1, \dots, m$, $j = 1, \dots, n_i$, where \mathbf{X}_{ij} denotes the covariates associated with observation j in cluster i , $\boldsymbol{\beta}_i$ is the cluster-specific parameter vector for cluster i , $f(\cdot)$ is a function known up to $\boldsymbol{\beta}_i$, and ϵ_{ij} is the intra-cluster random errors. Define $\mathbf{X}_i = (\mathbf{X}_{i1}^T, \dots, \mathbf{X}_{in_i}^T)^T$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T$, and assume $\boldsymbol{\epsilon}_i | \boldsymbol{\beta}_i \sim N\{\mathbf{0}, \mathbf{R}_i(\boldsymbol{\beta}_i, \boldsymbol{\xi})\}$, where $\boldsymbol{\xi}$ denotes parameters in the variance–covariance matrix of $\boldsymbol{\epsilon}_i$ besides $\boldsymbol{\beta}_i$. The

inter-cluster model is assumed to be

$$\beta_i = g(\mathbf{W}_i, \beta, \mathbf{b}_i), \quad i = 1, \dots, m, \tag{6}$$

where β is the vector of the fixed effects, \mathbf{b}_i are the random effects associated with cluster i , and \mathbf{W}_i denotes the observable attributes of cluster i . It is assumed that $\{Y_{ij}, j = 1, \dots, n_i\}$ are conditionally independent given $(\mathbf{X}_i, \mathbf{b}_i)$, $i = 1, \dots, m$.

Define $f_{ij} = f(\mathbf{X}_{ij}, \beta_i)$ and let $\mathbf{f}_i = (f_{i1}, \dots, f_{in_i})^T$, for $i = 1, \dots, m$. The density of the observed data from cluster i is then

$$f_{Y_i}(\mathbf{Y}_i; \Omega) = \int f_{\mathbf{b}_i}(\mathbf{b}_i; \tau) \frac{|\mathbf{R}_i(\beta_i, \xi)|^{-1/2}}{(2\pi)^{n_i/2}} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_i - \mathbf{f}_i)^T \mathbf{R}_i(\beta_i, \xi)^{-1} (\mathbf{Y}_i - \mathbf{f}_i) \right\} d\mathbf{b}_i, \tag{7}$$

where $f_{\mathbf{b}_i}(\mathbf{b}_i; \tau)$ is the assumed density of \mathbf{b}_i indexed by variance components τ , and $\Omega = (\beta^T, \xi^T, \tau^T)^T$. In this article, we assume that the nonlinear function $f(\cdot)$ and the intra-cluster variance–covariance matrix $\mathbf{R}_i(\cdot)$ are correctly specified but that $g(\cdot)$ in (6) and $f_{\mathbf{b}_i}(\mathbf{b}_i; \tau)$ in (7) may be incorrect.

3.2. Coarsening mechanism

There are more choices of coarsening mechanism for non-binary response than for binary response. From a practical standpoint, because the diagnostic method is likelihood-based, it is preferable that the chosen mechanism yields a tractable coarsened-data density. For illustrative purpose, we focus on two coarsening mechanisms that yield tractable likelihood in the forthcoming simulation study.

Both mechanisms start with partitioning cluster i into G_i groups, each of size k_{ig} , for $i = 1, \dots, m, g = 1, \dots, G_i$. According to the first mechanism, a new response for group g in cluster i is defined by

$$Y_{ig}^{(a)} = k_{ig}^{-1} \sum_{j \in I_{ig}} Y_{ij}, \quad i = 1, \dots, m, g = 1, \dots, G_i, \tag{8}$$

where I_{ig} denotes the index set corresponding to the observations in group g of cluster i . The second mechanism defines the new response as

$$Y_{ig}^{(b)} = \max_{j \in I_{ig}} Y_{ij}, \quad i = 1, \dots, m, g = 1, \dots, G_i. \tag{9}$$

We refer to the coarsened data with response defined in (8) as average-coarsened data (ACD), and to the coarsened data with response defined in (9) as maximum-coarsened data (MCD). Define $\mathbf{Y}_i^{(a)} = (Y_{i1}^{(a)}, \dots, Y_{iG_i}^{(a)})^T$ and $\mathbf{Y}_i^{(b)} = (Y_{i1}^{(b)}, \dots, Y_{iG_i}^{(b)})^T$, for $i = 1, \dots, m$. It can be shown that the density of $\mathbf{Y}_i^{(a)}$ is

$$f_{\mathbf{Y}_i^{(a)}}(\mathbf{Y}_i^{(a)}; \Omega) = \int f_{\mathbf{b}_i}(\mathbf{b}_i; \tau) \frac{|\mathbf{R}_i^*(\beta_i, \xi)|^{-1/2}}{(2\pi)^{G_i/2}} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_i^{(a)} - \mathbf{f}_i^*)^T \mathbf{R}_i^*(\beta_i, \xi)^{-1} (\mathbf{Y}_i^{(a)} - \mathbf{f}_i^*) \right\} d\mathbf{b}_i, \tag{10}$$

where $\mathbf{f}_i^* = \mathbf{S}_i \mathbf{f}_i$, $\mathbf{R}_i^*(\beta_i, \xi) = \mathbf{S}_i \mathbf{R}_i(\beta_i, \xi) \mathbf{S}_i^T$, and \mathbf{S}_i satisfies $\mathbf{Y}_i^{(a)} = \mathbf{S}_i \mathbf{Y}_i$. The density of $\mathbf{Y}_i^{(b)}$ is more involved. In a simple case with $\mathbf{R}_i(\beta_i, \xi) = \text{diag}(\sigma_{i1}^2, \dots, \sigma_{in_i}^2)$, the density of $\mathbf{Y}_i^{(b)}$ is given by

$$f_{\mathbf{Y}_i^{(b)}}(\mathbf{Y}_i^{(b)}; \Omega) = \int f_{\mathbf{b}_i}(\mathbf{b}_i; \tau) \prod_{g=1}^{G_i} \left[\prod_{j \in I_{ig}} \sigma_{ij}^{-1} \Phi \left\{ \sigma_{ij}^{-1} (Y_{ig}^{(b)} - f_{ij}) \right\} \right. \\ \left. \times \sum_{j \in I_{ig}} \phi \left\{ \sigma_{ij}^{-1} (Y_{ig}^{(b)} - f_{ij}) \right\} \Phi^{-1} \left\{ \sigma_{ij}^{-1} (Y_{ig}^{(b)} - f_{ij}) \right\} \right] d\mathbf{b}_i, \tag{11}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Denote by $\hat{\Omega}_c^{(a)}$ and $\hat{\Omega}_c^{(b)}$ the MLEs for Ω based on ACD and MCD, respectively, and by $\tilde{\Omega}_c^{(a)}$ and $\tilde{\Omega}_c^{(b)}$ their limits as $m \rightarrow \infty$ with $\max_{1 \leq i \leq m} G_i$ bounded. With the likelihood function well defined, the test statistics \mathbf{t}_1 and F_1 can be easily calculated and used to assess the sensitivity of MLEs to data coarsening under the presumed NLMM for non-binary response.

The power of the tests depends on how sensitive the inference is to the interaction of model assumptions and data coarsening. We now construct a measure of this sensitivity, which sheds some light on the nature of \mathbf{t}_1 and F_1 from a different angle. The proposed measure is closely related to a sensitivity measure defined by Gustafson (2001). He defined a ratio of two Kullback–Leibler (KL) divergence, with the numerator being the KL divergence of the asymptotic distribution of the misspecified-model MLE from the asymptotic distribution of the true-model MLE, and the denominator being the KL divergence of the misspecified model from the true model. For two distributions with densities, h_1 and h_2 , the KL divergence is defined by $\text{KL}(h_1 \parallel h_2) = \int \log(h_1/h_2) dH_1$, where H_1 is the distribution function corresponding to h_1 . Even though the true model is unknown in practice, Gustafson’s sensitivity measure can enhance one’s understanding of the effects of

model misspecification on likelihood inference. Following his idea, for a parameter γ , we use the following ratio of two KL divergence to assess the sensitivity of inference for γ ,

$$\frac{\text{KL}\{h(\hat{\gamma}) \parallel h(\hat{\gamma}_c)\}}{\text{KL}\{h^*(\hat{\gamma}) \parallel h^*(\hat{\gamma}_c)\}}, \quad (12)$$

where $h(\hat{\gamma})$ and $h(\hat{\gamma}_c)$ are the densities of the asymptotic distributions of $\hat{\gamma}$ and $\hat{\gamma}_c$ under the assumed model based on the observed data and the coarsened data, respectively, and $h^*(\hat{\gamma})$ and $h^*(\hat{\gamma}_c)$ are the counterpart densities under the true model. The numerator in (12) quantifies the discrepancy between the inference resulting from the coarsened data and that from the observed data based on the assumed model. Assuming $\hat{\gamma} \sim AN(\tilde{\gamma}, m^{-1}\zeta)$ and $\hat{\gamma}_c \sim AN(\tilde{\gamma}_c, m^{-1}\zeta_c)$, one can show that

$$\text{KL}\{h(\hat{\gamma}) \parallel h(\hat{\gamma}_c)\} = \frac{m(\tilde{\gamma}_c - \tilde{\gamma})^2}{2\zeta_c} + \frac{1}{2} \left\{ \frac{\zeta}{\zeta_c} - \log\left(\frac{\zeta}{\zeta_c}\right) - 1 \right\}. \quad (13)$$

Under the true model, since both MLEs are consistent estimators, one can assume that the asymptotic distributions of $\hat{\gamma}$ and $\hat{\gamma}_c$ are $AN(\gamma_0, m^{-1}\zeta^*)$ and $AN(\gamma_0, m^{-1}\zeta_c^*)$, respectively, where γ_0 denotes the true parameter value. Consequently, the denominator of (12) equals

$$\text{KL}\{h^*(\hat{\gamma}) \parallel h^*(\hat{\gamma}_c)\} = \frac{1}{2} \left\{ \frac{\zeta^*}{\zeta_c^*} - \log\left(\frac{\zeta^*}{\zeta_c^*}\right) - 1 \right\}, \quad (14)$$

which measures the discrepancy in inference for γ based on the observed data and the coarsened data in the absence of model misspecification. The ratio of (13) and (14) can be viewed as a standardized divergence that assesses the difference between the inference for γ using the coarsened data and the inference based on the observed data after adjusting for the inevitable efficiency loss due to data coarsening. Clearly, (12) equals one when the assumed model coincides with the true model. When the true model is known as in a simulation setting, this ratio can be used to explicitly investigate the effect of using different coarsened data on the subsequent inference when model assumptions are violated. In practice, when the true model is unknown, and so is (14), one can consider the leading term in (12) as $m \rightarrow \infty$ to compare the sensitivity of different coarsened data to model misspecification, that is,

$$\frac{(\tilde{\gamma}_c - \tilde{\gamma})^2}{\zeta_c \left\{ \frac{\zeta^*}{\zeta_c^*} - \log\left(\frac{\zeta^*}{\zeta_c^*}\right) - 1 \right\}} \propto \frac{(\tilde{\gamma}_c - \tilde{\gamma})^2}{\zeta_c}. \quad (15)$$

In fact, one can construct a statistic to diagnose model misspecification based on the second expression in (15) provided that one can derive the null distribution of this statistic. Indeed, the second expression in (15) is akin in spirit to \mathbf{t}_1 and F_1 , whose null distributions are in familiar distribution families. This similarity hints at the connection between the proposed diagnostic tests and the measure of sensitivity of coarsened data to model misspecification originating from the KL divergence.

4. Implementation of diagnostic methods

4.1. New test statistics

To lessen computational burden, we construct a new set of test statistics that is free of $\hat{\Omega}_c$. This feature is especially appealing when fitting the model for the coarsened data is more numerically problematic than fitting the model for the observed data. The motivation of these test statistics is that, under H_0 , because both $\hat{\Omega}$ and $\hat{\Omega}_c$ are consistent, the solution to the normal score equation associated with the observed data should approximately solve the normal score equation for the coarsened data. Based on this logic, we define the following test statistic,

$$\mathbf{t}_2 = \left\{ m^{-1/2} \sum_{i=1}^m \Psi_c(\mathbf{Y}_i^*; \hat{\Omega}) \right\} \# \text{vecdiag}(\hat{\mathbf{V}}_2^{-1}), \quad (16)$$

where \mathbf{Y}_i^* is the coarsened data for cluster i , $\Psi_c(\mathbf{Y}_i^*; \hat{\Omega})$ is the partial derivative of the log likelihood of \mathbf{Y}_i^* with respect to Ω evaluated at $\hat{\Omega}$, and $\hat{\mathbf{V}}_2$ is an estimator for the variance-covariance matrix of $m^{-1/2} \sum_{i=1}^m \Psi_c(\mathbf{Y}_i^*; \hat{\Omega})$ (derived in the Appendix). Denote by $t_2(\gamma)$ the element in \mathbf{t}_2 corresponding to parameter γ . Under H_0 , $t_2(\gamma)$ is close to zero. A value of $t_2(\gamma)$ that significantly deviates from zero implies possible model misspecification. It is shown in the Appendix that $t_2(\gamma)$ follows a t distribution asymptotically with $m - p$ degrees of freedom under H_0 . Analogous to the construction of F_1 in (2), we define a test statistic based on \mathbf{t}_2 that combines the elementwise evidence as follows,

$$F_2 = \frac{m - p}{mp(m - 1)} \left\{ \sum_{i=1}^m \Psi_c(\mathbf{Y}_i^*; \hat{\Omega}) \right\}^T \hat{\mathbf{V}}_2^{-1} \left\{ \sum_{i=1}^m \Psi_c(\mathbf{Y}_i^*; \hat{\Omega}) \right\}, \quad (17)$$

Table 1

Monte Carlo averages of maximum likelihood estimates for Ω across 200 replicated data sets in Example 1 in Section 4.2. Numbers in parentheses are the Monte Carlo standard errors. The true parameter values are $\beta_1 = 1, \beta_2 = 5, \beta_3 = 2, \sigma^2 = 1$, and $\sigma_1^2 = 1.5$.

| | Observed data | ACD | MCD | Observed data | ACD | MCD |
|--------------------|---|--------------|--------------|---|--------------|--------------|
| | [Correct model] | | | [(C1) with $\theta = 0$] | | |
| $\hat{\beta}_1$ | 1.00 (0.005) | 1.00 (0.005) | 1.00 (0.006) | 1.00 (0.005) | 1.00 (0.005) | 1.01 (0.007) |
| $\hat{\beta}_2$ | 4.99 (0.013) | 4.99 (0.015) | 5.00 (0.026) | 4.99 (0.014) | 5.00 (0.017) | 5.06 (0.032) |
| $\hat{\beta}_3$ | 2.00 (0.012) | 1.99 (0.014) | 2.00 (0.020) | 1.99 (0.011) | 2.00 (0.014) | 2.03 (0.025) |
| $\hat{\sigma}^2$ | 1.00 (0.001) | 0.99 (0.003) | 1.00 (0.003) | 1.00 (0.001) | 0.99 (0.003) | 1.00 (0.003) |
| $\hat{\sigma}_1^2$ | 1.51 (0.011) | 1.52 (0.013) | 1.52 (0.016) | 1.50 (0.009) | 1.52 (0.011) | 1.53 (0.016) |
| | [(C1) with $\theta = 0.5$] | | | [(C2) missing b_{i2} , with $\theta = 0.5$] | | |
| $\hat{\beta}_1$ | 0.79 (0.013) | 1.06 (0.007) | 0.70 (0.008) | 0.76 (0.014) | 1.15 (0.010) | 0.93 (0.007) |
| $\hat{\beta}_2$ | 4.87 (0.010) | 5.09 (0.022) | 4.89 (0.019) | 5.34 (0.012) | 5.46 (0.023) | 5.73 (0.022) |
| $\hat{\beta}_3$ | 2.04 (0.005) | 1.99 (0.013) | 2.06 (0.012) | 2.13 (0.005) | 2.13 (0.012) | 2.19 (0.011) |
| $\hat{\sigma}^2$ | 1.47 (0.012) | 0.36 (0.002) | 2.05 (0.043) | 1.17 (0.005) | 0.29 (0.001) | 1.24 (0.010) |
| $\hat{\sigma}_1^2$ | 0.26 (0.007) | 1.87 (0.030) | 0.42 (0.013) | 0.59 (0.011) | 3.23 (0.045) | 1.04 (0.018) |
| $\hat{\theta}$ | 1.07 (0.003) | 0.95 (0.004) | 1.07 (0.005) | 0.98 (0.003) | 0.83 (0.004) | 0.90 (0.004) |
| | [(C3) missing b_{i2} and b_{i3} , with $\sigma_3^2 = 1.5$] | | | [(C3) missing b_{i2} and b_{i3} , with $\sigma_3^2 = 2$] | | |
| $\hat{\beta}_1$ | 0.92 (0.005) | 0.93 (0.005) | 0.83 (0.006) | 0.89 (0.005) | 0.90 (0.004) | 0.77 (0.006) |
| $\hat{\beta}_2$ | 4.72 (0.018) | 4.69 (0.016) | 4.74 (0.025) | 4.63 (0.019) | 4.55 (0.019) | 4.66 (0.027) |
| $\hat{\beta}_3$ | 1.96 (0.014) | 1.94 (0.017) | 1.76 (0.022) | 2.04 (0.016) | 2.01 (0.018) | 1.73 (0.020) |
| $\hat{\sigma}^2$ | 1.07 (0.002) | 1.28 (0.007) | 1.15 (0.005) | 1.09 (0.002) | 1.39 (0.007) | 1.19 (0.005) |
| $\hat{\sigma}_1^2$ | 1.29 (0.010) | 1.04 (0.010) | 1.15 (0.013) | 1.22 (0.010) | 0.90 (0.009) | 1.05 (0.012) |

which follows an $F(p, m - p)$ distribution asymptotically under H_0 by the construction of $\hat{\mathbf{V}}_2$ given in the Appendix. Similar to \mathbf{t}_1 and F_1 , the pattern of the individual test statistics \mathbf{t}_2 is related to the source of model misspecification, and F_2 gives an overall assessment of the sensitivity of inference to data coarsening under the current model assumptions.

4.2. Simulation evidence

Two examples in NLMM are used next to illustrate the implementation and performance of the proposed tests.

Example 1 (Growth Model). Mimicking the logistic model for the growth of orange trees in Pinheiro and Bates (1995), one posits an NLMM with intra-cluster model

$$Y_{ij} = \frac{\beta_{i1}}{1 + \exp\{-(t_{ij} - \beta_{2i})/\beta_{3i}\}} + \epsilon_{ij}, \quad i = 1, \dots, m, j = 1, \dots, n_i, \tag{18}$$

where $\beta_i = (\beta_{i1}, \beta_{i2}, \beta_{i3})^T$ is the vector of cluster-specific parameters, t_{ij} is the observed covariate representing time, $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T \sim N\{0, \mathbf{R}_i(\beta_i, \xi)\}$, $\mathbf{R}_i(\beta_i, \xi) = \sigma^2 \text{diag}(f_{i1}^{2\theta}, \dots, f_{in_i}^{2\theta})$, and $\xi = (\sigma^2, \theta)^T$. One further assumes the inter-cluster model to be

$$\beta_{i1} = \beta_1 + b_{i1}, \quad \beta_{i2} = \beta_2, \quad \beta_{i3} = \beta_3, \quad i = 1, \dots, m, \tag{19}$$

where $\beta = (\beta_1, \beta_2, \beta_3)^T$ is the vector of fixed effects and the random effect $b_{i1} \sim N(0, \sigma^2 \sigma_1^2)$.

Before considering model misspecification, we first check the size of the tests. We generate 200 Monte Carlo data sets from the model specified in (18) and (19), with $m = 400, n_i = 12, t_{ij} = j$, for $i = 1, \dots, 400, j = 1, \dots, 12$, and $\theta = 0$ so that $\xi = \sigma^2$. The true parameter values are $\beta_1 = 1, \beta_2 = 5, \beta_3 = 2, \sigma^2 = 1$, and $\sigma_1^2 = 1.5$. For the coarsened data, we form ACD and MCD by first partitioning each cluster into three groups of equal size, yielding $G_i = 3$ and $k_{ig} = 4$, for $i = 1, \dots, 400, g = 1, 2, 3$. In order to reduce information loss during coarsening, each cluster is partitioned according to the sorted time points so that the ratio of the between-group variance of the time points over the within-group variance of the time points is maximized. Under the assumed model, the closed-form expressions of (7) and (10), that is, the densities of \mathbf{Y}_i and $\mathbf{Y}_i^{(a)}$, can be easily derived. Closed-form expression for the density of $\mathbf{Y}_i^{(b)}$ in (11) is generally not available. In the simulation presented in this subsection, we use 50-point Gauss–Hermite quadrature to approximate the integral in (11).

For each Monte Carlo replicate, we compute $\hat{\Omega}, \hat{\Omega}_c^{(a)}, \hat{\Omega}_c^{(b)}$, and the test statistics, $\mathbf{t}_1, F_1, \mathbf{t}_2$, and F_2 . The Monte Carlo averages of the estimates from this simulation are summarized in the upper-left block of Table 1. The proportion of Monte Carlo replicates that reject H_0 for each test statistic is presented in the top portion of Table 2. As expected, when the model for the observed data is correct, none of the three sets of MLEs show noticeable bias. With all rejection rates around the nominal 0.05 significance level, these results suggest that the test statistics confer the correct size.

Table 2

Rejection rates of the test statistics across 200 Monte Carlo replicates in Example 1 in Section 4.2. Rejection rates higher than 0.30 are in italic. The most significant tests for the fixed effects in the two cases with true models being heteroscedastic mixed models are underlined.

| | t_1 | | | | | F_1 | t_2 | | | | | F_2 |
|---|-------------|-------------|-----------|------------|---------------|--------|-------------|-------------|-----------|------------|---------------|--------|
| | $(\beta_1$ | β_2 | β_3 | σ^2 | $\sigma_1^2)$ | | $(\beta_1$ | β_2 | β_3 | σ^2 | $\sigma_1^2)$ | |
| [Correct model] | | | | | | | | | | | | |
| ACD | 0.05 | 0.05 | 0.06 | 0.04 | 0.04 | {0.04} | 0.07 | 0.05 | 0.03 | 0.04 | 0.04 | {0.03} |
| MCD | 0.03 | 0.04 | 0.04 | 0.07 | 0.05 | {0.05} | 0.04 | 0.05 | 0.03 | 0.07 | 0.06 | {0.06} |
| [(C1) with $\theta = 0$] | | | | | | | | | | | | |
| ACD | 0.04 | 0.05 | 0.04 | 0.06 | 0.05 | {0.04} | 0.06 | 0.05 | 0.05 | 0.06 | 0.06 | {0.07} |
| MCD | 0.04 | 0.02 | 0.03 | 0.05 | 0.03 | {0.06} | 0.07 | 0.06 | 0.07 | 0.03 | 0.04 | {0.06} |
| [(C1) with $\theta = 0.5$, heteroscedastic mixed model] | | | | | | | | | | | | |
| ACD | <u>0.72</u> | 0.06 | 0.08 | 0.99 | 0.98 | {1.00} | <u>0.58</u> | 0.09 | 0.16 | 0.40 | 0.03 | {1.00} |
| MCD | <u>0.32</u> | 0.05 | 0.03 | 0.27 | 0.27 | {1.00} | <u>0.66</u> | 0.24 | 0.40 | 0.51 | 0.40 | {1.00} |
| [(C2) missing b_{i2} , heteroscedastic mixed model] | | | | | | | | | | | | |
| ACD | 0.03 | <u>0.58</u> | 0.02 | 1.00 | 0.99 | {1.00} | 0.04 | <u>0.41</u> | 0.08 | 0.69 | 0.10 | {1.00} |
| MCD | 0.13 | <u>0.22</u> | 0.04 | 0.09 | 0.51 | {0.98} | 0.21 | <u>0.57</u> | 0.05 | 0.73 | 0.50 | {0.98} |
| [(C3) missing b_{i2} and b_{i3} , with $\sigma_3^2 = 1.5$] | | | | | | | | | | | | |
| ACD | 0.08 | 0.09 | 0.05 | 0.86 | 0.88 | {0.78} | 0.07 | 0.09 | 0.05 | 0.88 | 0.07 | {0.66} |
| MCD | 0.44 | 0.04 | 0.22 | 0.37 | 0.31 | {0.37} | 0.49 | 0.10 | 0.08 | 0.28 | 0.08 | {0.36} |
| [(C3) missing b_{i2} and b_{i3} , with $\sigma_3^2 = 2$] | | | | | | | | | | | | |
| ACD | 0.07 | 0.14 | 0.06 | 0.99 | 0.99 | {0.96} | 0.08 | 0.09 | 0.05 | 1.00 | 0.05 | {0.92} |
| MCD | 0.68 | 0.03 | 0.24 | 0.47 | 0.42 | {0.61} | 0.76 | 0.10 | 0.08 | 0.36 | 0.06 | {0.59} |

Next we empirically investigate the power of the tests. Simulated data sets are generated from the following three (true) models.

- (C1) The true inter-cluster model is as in (19), with $b_{i1} \sim 0.2N(2, 0.4\sigma^2) + 0.8N(-0.5, 0.54\sigma^2)$, yielding $\text{var}(b_{i1}) = 1.5\sigma^2$. In the intra-cluster model that is common for the true model and the assumed model, we first set $\theta = 0$, resulting in a homoscedastic mixed model; then, we raise θ to 0.5, producing a heteroscedastic mixed model.
- (C2) Set $\theta = 0.5$ for both assumed and true models. The true inter-cluster model is

$$\beta_{1i} = \beta_1 + b_{i1}, \quad \beta_{2i} = \beta_2 + b_{i2}, \quad \beta_{3i} = \beta_3, \quad i = 1, \dots, m, \tag{20}$$

where $\mathbf{b}_i = (b_{i1}, b_{i2})^T \sim N(0, \sigma^2 \Sigma_b)$, and the elements in Σ_b are $\Sigma_b[1, 1] = \sigma_1^2 = 1.5$, $\Sigma_b[2, 2] = \sigma_2^2 = 1$, and $\Sigma_b[1, 2] = \Sigma_b[2, 1] = 0.8$.

- (C3) Set $\theta = 0$ for both assumed and true models. The true inter-cluster model is

$$\beta_{1i} = \beta_1 + b_{i1}, \quad \beta_{2i} = \beta_2 + b_{i2}, \quad \beta_{3i} = \beta_3 + b_{i3}, \quad i = 1, \dots, m, \tag{21}$$

where $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})^T \sim N\{0, \sigma^2 \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)\}$, with $\sigma_1^2 = 1.5$, $\sigma_2^2 = 1$, and $\sigma_3^2 = 1.5$ or 2. The true model with $\sigma_3^2 = 2$ represents a further deviation of the assumed model from the true model than when $\sigma_3^2 = 1.5$.

The estimates, $\hat{\Omega}$, $\hat{\Omega}_c^{(a)}$, $\hat{\Omega}_c^{(b)}$, and the test statistics are computed using the simulated data from (C1)–(C3), respectively, based on the assumed model defined in (18) and (19). The Monte Carlo averages of the estimates and the rejection rates of the test statistics are tabulated in Tables 1 and 2.

Under (C1), the only assumption violated in the assumed model is the distribution for the random effect b_{i1} . From Table 1, when $\theta = 0$, that is, the true model, as the assumed model, is a homoscedastic NLMM, the MLEs are robust to the distributional assumption on the only random effect b_{i1} . This is a reminiscence of the robustness of MLEs in LMM. Verbeke and Lesaffre (1997) showed that MLEs for the fixed effects and variance components in LMM are consistent even when the random-effects distribution is misspecified. In fact, the proof for the consistency of MLE in homoscedastic LMM in Verbeke and Lesaffre (1994) also applies to homoscedastic NLMM when the response is linear in the random effects, as in (C1). Therefore, the robustness evidence observed in (C1) with $\theta = 0$ is expected. The rejection rates in Table 2 under (C1) with $\theta = 0$ reinforce that the size of the test statistics is well controlled. However, the proof in Verbeke and Lesaffre (1994) does not carry over to heteroscedastic mixed models, even when the response is linear in the random effects, as in (C1) with $\theta = 0.5$. There, the deterioration of the MLEs is clearly shown in Table 1. The corresponding test statistics in Table 2 show promising power. Most notably, the tests for β_1 , with which the misspecified random effect b_{i1} is associated, tend to be more significant than the tests for the other fixed effects. This phenomenon bears an obvious resemblance to the previous finding in GLMM that the pattern of the individual tests is related to the source of misspecification.

The intriguing dependence of the individual tests on the source of model misspecification is also present under (C2). There, the assumed model excludes the random effect associated with β_2 . As a result, $t_1(\beta_2)$ and $t_2(\beta_2)$ stand out as being

Table 3

Rejection rates of the test statistics across 300 Monte Carlo replicates in Example 2 in Section 4.2. Rejection rates higher than 0.30 are in italic.

| | (D1) | | (D2) | | (D3) | |
|-------------------|-------------|-------------|-------------|-------------|-------|-------|
| | G = 3 | G = 4 | G = 3 | G = 4 | G = 3 | G = 4 |
| $t_1(\beta_1)$ | 0.07 | 0.06 | <i>0.89</i> | <i>0.92</i> | 0.09 | 0.04 |
| $t_1(\beta_2)$ | 0.06 | 0.09 | 0.05 | 0.04 | 0.08 | 0.06 |
| $t_1(\beta_3)$ | 0.05 | 0.06 | <i>0.89</i> | <i>0.87</i> | 0.08 | 0.04 |
| $t_1(\sigma^2)$ | <i>0.68</i> | <i>0.58</i> | 0.09 | 0.05 | 1 | 1 |
| $t_1(\sigma_1^2)$ | <i>0.77</i> | <i>0.63</i> | 0.07 | 0.05 | 1 | 1 |
| $t_1(\sigma_2^2)$ | <i>0.74</i> | <i>0.60</i> | 0.06 | 0.05 | (NA) | (NA) |
| $t_1(\rho)$ | 0.04 | 0.06 | 0.07 | 0.08 | (NA) | (NA) |
| F_1 | <i>0.65</i> | <i>0.42</i> | 1 | 1 | 1 | 1 |

most significant among the tests for the fixed effects. This association between the individual test statistics and the source of model misspecification is not evident for homoscedastic mixed models such as (C3). However, results from (C3) clearly demonstrate that the magnitude of the bias in most of MLEs increases as the assumed model deviates further from the true model (with σ_3^2 increasing from 1.5 to 2). Both overall tests based on F_1 and F_2 have moderate to high power, so do many of the individual tests. The power increases as misspecification becomes more severe.

Example 2 (One-Compartment Model). A one-compartment model, considered in Hartford and Davidian (2000), is suitable in characterizing the plasma concentration of a subject after the subject is given a drug as an intravenous bolus. Suppose the true intra-subject model is $Y_{ij} = D_i U_i^{-1} \exp(-C_i t_{ij} U_i^{-1}) + \epsilon_{ij}$, where t_{ij} is the time when the j th measurement of plasma concentration for subject i is taken, D_i denotes the dose for subject i , C_i represents the clearing distribution of subject i , U_i is the subject-specific volume of distribution, and $\epsilon_i \sim N(0, \mathbf{R}_i(\boldsymbol{\beta}_i, \sigma^2))$, for $i = 1, \dots, m, j = 1, \dots, n_i$. The true inter-subject model for $\boldsymbol{\beta}_i = (C_i, U_i)^T$ is

$$C_i = \exp\left(\beta_1 + \frac{a_i}{100}\beta_2 + b_{i1}\right), \quad U_i = \exp(\beta_3 + b_{i2}), \quad i = 1, \dots, m, \tag{22}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ is the vector of fixed effects, $\mathbf{b}_i = (b_{i1}, b_{i2})^T$ is the vector of random effects, and a_i is a covariate associated with subject i .

We use three simulation settings as follows.

(D1) In the true model, \mathbf{b}_i follows a bivariate mixture normal distribution,

$$0.9N\left\{\begin{pmatrix} 0.01 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 0.046 & 0.052 \\ 0.052 & 0.086 \end{pmatrix}\right\} + 0.1N\left\{\begin{pmatrix} -0.09 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 0.046 & 0.052 \\ 0.052 & 0.086 \end{pmatrix}\right\}.$$

This yields $\text{var}(b_i)$ equal to

$$\sigma^2 \begin{pmatrix} 0.046 & 0.052 \\ 0.052 & 0.086 \end{pmatrix}. \tag{23}$$

In the assumed model, b_i follows a bivariate normal distribution. In both true and assumed models, $\mathbf{R}_i(\boldsymbol{\beta}_i, \sigma^2) = \sigma^2 \mathbf{I}_{n_i}$.

(D2) The true model and the assumed model are identical to those in (D1) except that $\mathbf{R}_i(\boldsymbol{\beta}_i, \sigma^2) = \text{diag}(\sigma^2 f_{i1}^2, \dots, \sigma^2 f_{in_i}^2)$.

(D3) In the true model, $b_i \sim N(0, \sigma^2 \Sigma_b)$, where $\sigma^2 \Sigma_b$ is equal to (23). In the assumed model, b_{i2} in (22) is excluded and $b_{i1} \sim N(0, \sigma^2 \sigma_1^2)$.

Common settings in (D1)–(D3) are $m = 100, \beta_1 = 0.223, \beta_2 = 0.6, \beta_3 = 2.303, \sigma^2 = 0.023, D_i = 100, n_i = 12$, and $t_i = (0.25, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 24)^T$, and a_i is generated from $N(70, 20^2)$, for $i = 1, \dots, 100$. The number of Monte Carlo replications is 300.

In this simulation, we focus on ACD / coarsened data, where each cluster is partitioned into $G_i = G (=3 \text{ or } 4)$ groups of equal size, $i = 1, \dots, 100$. We vary the value of G to explore the effect of group size on the tests. Closed-form expression for the integrals in (7) and (10) are not available. We employ Laplacian approximation (Pinheiro and Bates, 1995; Hartford and Davidian, 2000) to compute the integrals. The rejection rates of the first set of test statistics are summarized in Table 3.

The MLEs (not tabulated due to space limitation) for the fixed effects are not noticeably affected when the model is homoscedastic, as in (D1) and (D3), but those for the variance components are severely biased. The low rejection rates associated with the tests for the fixed effects and the high rejection rates associated with those for the variance components for these two cases are in agreement with these findings. But for a heteroscedastic model, as in (D2), the MLEs for β_1 and β_2 are obviously biased and the test statistics associated with them also suggest the existence of bias. The bias and the rejection rates also depend on G . These findings concur with those in Section 2.2 for GLMM, where it is demonstrated that the bias in MLEs can depend on both G and the source of misspecification.

Because NLMM is such a rich class of models and there are many factors contributing to the hierarchical modeling, it is impossible to summarize the effect of different types of model misspecification on MLEs for different parameters. Our hope is

Table 4

Maximum likelihood estimates for the parameters in the assumed models for soybean growth data. Numbers in parentheses are the estimated standard errors.

| Assume a normal random effect b_{i1} | | | | | | | | |
|--|-----------------|------------------|------------------|--------------------|-----------------|-----------------|------------------|-----------------|
| $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ | $\hat{\sigma}_1^2$ | $\hat{\theta}$ | $\hat{\alpha}$ | | |
| Observed data | | | | | | | | |
| 17.66 (0.421) | 5.19 (0.059) | -1.33 (0.023) | 0.06 (0.006) | 20.06 (1.638) | 0.94 (0.023) | 0.22 (0.046) | | |
| ACD | | | | | | | | |
| 18.07 (0.475) | 5.27 (0.069) | -1.28 (0.028) | 0.06 (0.010) | 20.62 (2.083) | 0.95 (0.055) | 0.22 (0.125) | | |
| Assume a mixture normal random effect b_{i1} | | | | | | | | |
| $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ | $\hat{\sigma}_1^2$ | $\hat{\theta}$ | $\hat{\alpha}$ | $\hat{\mu}_{11}$ | $\hat{\delta}$ |
| Observed data | | | | | | | | |
| 17.47 (0.062) | 5.19 (0.062) | -1.32 (0.024) | 0.06 (0.006) | 3.99 (1.472) | 0.92 (0.026) | 0.22 (0.042) | 7.62 (0.721) | 0.18 (0.057) |
| ACD | | | | | | | | |
| 17.93 (0.067) | 5.30 (0.076) | -1.27 (0.033) | 0.07 (0.014) | 2.98 (1.685) | 0.88 (0.062) | 0.26 (0.100) | 7.64 (0.742) | 0.17 (0.056) |

to detect the existence of misspecification and narrow down to the most likely source of misspecification using the proposed method. The reported simulation studies provide strong evidence of achieving this goal, especially for heteroscedastic mixed effects models where the individual tests show potential to identify the source of misspecification.

5. Application to soybean growth data

We now apply the diagnostic method to a soybean growth data set in Davidian and Giltinan (1995, Chapter 11, Section 11.2). In this study, from each of $m = 48$ plots of soybean, the average leaf weight per plant is collected at different time points after planting. Let Y_{ij} denote the average leaf weight per plant in plot i at time t_{ij} , for $i = 1, \dots, 48$ and $j = 1, \dots, n_i$, where n_i varies between 8 and 10. We posit the following growth model,

$$Y_{ij} = \frac{\beta_{i1}}{1 + \exp\{\beta_{i3}(t_{ij}/10 - \beta_{2i})\}} + \epsilon_{ij}, \quad i = 1, \dots, m, j = 1, \dots, n_i,$$

where $\beta_i = (\beta_{i1}, \beta_{i2}, \beta_{i3})^T$ is the vector of plot-specific parameters. Following Davidian and Giltinan (1995), the intra-plot variance-covariance structure is

$$R_i(\beta_i, \xi) = \sigma^2 L_i^{1/2}(\beta_i, \theta) \Gamma_i(\alpha) L_i^{1/2}(\beta_i, \theta),$$

where $L_i^{1/2}(\beta_i, \theta) = \text{diag}(|f_{i1}^\theta|, \dots, |f_{in_i}^\theta|)$, and

$$\Gamma_i(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n_i-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{n_i-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha^{n_i-1} & \alpha^{n_i-2} & \alpha^{n_i-3} & \dots & 1 \end{pmatrix}.$$

Unlike the analysis presented in Davidian and Giltinan (1995), which takes into account the soybean genotype and growing season, we aggregate these effects and consider only the growth pattern over time.

We apply the proposed diagnostic method to the assumed models with different inter-plot models, such as those in (19)–(21), with different distributional assumptions on the random effect(s). For the coarsened data, we use ACD with $G_i = 4$ or 5 to retain two to three observations per group per plot. For illustration purposes, we present the results for two relatively simple assumed models using the first set of test statistics. The simplest candidate model is (19), in which b_{i1} is assumed normal. The overall test in (2) gives $F_1 = 4.80$ with p -value 0.0005, providing strong evidence of model misspecification. The individual t tests associated with all three fixed effects are significant. We then consider revising the assumed model for b_{i1} to be a location mixture normal with density $\delta N(\mu_{11}, \sigma_1^2) + (1 - \delta)N\{-\mu_{11}\delta/(1 - \delta), \sigma_1^2\}$. For this assumed model, $F_1 = 1.97$ with p -value 0.07, indicating some evidence of model misspecification, but not enough at the 0.05 significance level. This indicates some improvement in the second assumed model over the first one. The parameter estimates for these two assumed models are given in Table 4. Note that, even though there are more parameters to estimate in the second assumed model than the first one, the estimated standard errors of $\hat{\Omega}$ in the second model are still mostly similar to those in the first model, except for the standard errors of $\hat{\beta}_1$ and $\hat{\sigma}_1^2$, which are much lower in the second model. This reconciles the general belief that model misspecification (in the first assumed model) can compromise efficiency. For

the second assumed model, we find $t_1(\beta_2)$ and $t_1(\beta_3)$ border-line significant at 0.05 level. But the results from individual tests need to be interpreted with cautious because using all individual tests simultaneously to test H_0 is a process of multiple testing. Therefore, the critical value for these individual tests should be adjusted accordingly to set family-wise type I error at 0.05. We do not pursue further the issue of multiple testing here, while acknowledging that there may be more appropriate assumed models for β_{2i} and β_{3i} than the current assumed models.

6. Discussion

We revisit the method for diagnosing random-effects model misspecification in GLMM for binary response proposed by Huang (2009) and further explore the theoretical driven force of the method. We then extend the method to a much richer class of mixed effects models for non-binary response and show that, using coarsened data, one can reveal random-effects model misspecification. For heteroscedastic mixed effects models, one may be able to identify the source of misspecification from the pattern of the individual test statistics. This is the main advantage of our method over most existing diagnostic methods and deserves further investigation in more general context of model selection.

We have focused on misspecification on the random-effects part of the models assuming absence of other source of misspecification in this study to avoid the untestable issue pointed out by Verbeke and Molenberghs (submitted for publication). It is of interest to investigate the operating characteristics of the proposed tests when other parts of the model are misspecified, for example, the fixed-effects part of the model or the intra-cluster variance-covariance structure. To develop diagnostic methods that can disentangle multiple sources of model misspecification in mixed effects models and pinpoint each source is a challenging avenue for future research.

Appendix. Derivation of variance estimator \hat{V}_2

Denote by $\psi(\mathbf{Y}_i; \Omega)$ the normal score function associated with the observed data in cluster i . Define

$$\mathbf{A}(\Omega) = \lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E \{ -\partial \psi(\mathbf{Y}_i; \Omega) / \partial \Omega^T \}, \quad \hat{\mathbf{A}}(\Omega) = -m^{-1} \sum_{i=1}^m \partial \psi(\mathbf{Y}_i; \Omega) / \partial \Omega^T,$$

$$\mathbf{A}_c(\Omega) = \lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E \{ -\partial \psi_c(\mathbf{Y}_i^*; \Omega) / \partial \Omega^T \}, \quad \hat{\mathbf{A}}_c(\Omega) = -m^{-1} \sum_{i=1}^m \partial \psi_c(\mathbf{Y}_i^*; \Omega) / \partial \Omega^T.$$

Under $H_0 : \tilde{\Omega} = \tilde{\Omega}_c$, a first-order Taylor expansion of $m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \hat{\Omega})$ around $\tilde{\Omega}$ yields

$$m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \hat{\Omega}) \tag{24}$$

$$\begin{aligned} &\simeq m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \tilde{\Omega}) + m^{-1/2} \sum_{i=1}^m \frac{\partial}{\partial \Omega^T} \psi_c(\mathbf{Y}_i^*; \Omega) \Big|_{\Omega=\tilde{\Omega}} (\hat{\Omega} - \tilde{\Omega}) \\ &\simeq m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \tilde{\Omega}) - m^{1/2} \mathbf{A}_c(\tilde{\Omega}) (\hat{\Omega} - \tilde{\Omega}) \end{aligned} \tag{25}$$

$$\simeq m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \tilde{\Omega}) - m^{-1/2} \mathbf{A}_c(\tilde{\Omega}) \mathbf{A}^{-1}(\tilde{\Omega}) \sum_{i=1}^m \psi(\mathbf{Y}_i; \tilde{\Omega}) \tag{26}$$

$$= m^{-1/2} \sum_{i=1}^m \{ \psi_c(\mathbf{Y}_i^*; \tilde{\Omega}) - \mathbf{A}_c(\tilde{\Omega}) \mathbf{A}^{-1}(\tilde{\Omega}) \psi(\mathbf{Y}_i; \tilde{\Omega}) \}, \tag{27}$$

where the approximation in (26) is based on an approximation of $\hat{\Omega} - \tilde{\Omega}$ in (25) with the influence function.

Denote by $\eta_i = \psi_c(\mathbf{Y}_i^*; \tilde{\Omega}) - \mathbf{A}_c(\tilde{\Omega}) \mathbf{A}^{-1}(\tilde{\Omega}) \psi(\mathbf{Y}_i; \tilde{\Omega})$. Based on the approximation in (27), a variance-covariance estimator for (24) is the sample variance-covariance of $\{\eta_i, i = 1, \dots, m\}$. Because η_i depends on the unknown $\tilde{\Omega}$, $\mathbf{A}(\tilde{\Omega})$, and $\mathbf{A}_c(\tilde{\Omega})$, we define an estimator for the variance-covariance matrix of (24), \hat{V}_2 , as the sample variance-covariance of $\{\eta_i, i = 1, \dots, m\}$ with $\tilde{\Omega}$ replaced by $\hat{\Omega}$, $\mathbf{A}(\tilde{\Omega})$ replaced by $\hat{\mathbf{A}}(\hat{\Omega})$, and $\mathbf{A}_c(\tilde{\Omega})$ replaced by $\hat{\mathbf{A}}_c(\hat{\Omega})$.

Now that under H_0 , (24) can be approximated by the sum of independent quantities, $m^{-1/2} \sum_{i=1}^m \eta_i$, of which the sample variance-covariance is estimated by \hat{V}_2 , by Corollary 3.5.1.1 in Mardia et al. (1979),

$$\left\{ m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \hat{\Omega}) \right\}^T \hat{V}_2^{-1} \left\{ m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \hat{\Omega}) \right\}$$

has the form of a Hotelling's T^2 statistic. It follows that

$$\frac{m-p}{p(m-1)} \left\{ m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \hat{\boldsymbol{\Omega}}) \right\}^T \hat{\mathbf{V}}_2^{-1} \left\{ m^{-1/2} \sum_{i=1}^m \psi_c(\mathbf{Y}_i^*; \hat{\boldsymbol{\Omega}}) \right\},$$

which is equal to F_2 in (17), follows $F(m, m-p)$ asymptotically under H_0 . By the construction of t_2 in (16), it is obvious that $t_2(\gamma)$ follows a Student's t distribution with $m-p$ degrees of freedom asymptotically under H_0 .

References

- Agresti, A., Caffo, B., 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.
- Barnhart, H.X., Williamson, J., 1998. Goodness-of-fit tests for GEE modeling with binary responses. *Biometrics* 54, 720–729.
- Davidian, M., Gallant, A.R., 1993. The nonlinear mixed effect model with a smooth random effects density. *Biometrika* 80, 475–488.
- Davidian, M., Giltinan, D.M., 1995. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC.
- Davidian, M., Giltinan, D.M., 2003. Nonlinear models for repeated measurement data: an overview and update. *Journal of Agricultural, Biological and Environmental Statistics* 8, 387–419.
- Fattinger, K.E., Sheiner, L.B., Verotta, D., 1995. A new method to explore the distribution of interindividual random effects in non-linear mixed effects models. *Biometrics* 51, 1236–1251.
- Gustafson, P., 2001. On measuring sensitivity to parametric model misspecification. *Journal of Royal Statistical Society, Series B* 63, 81–94.
- Hartford, A., Davidian, M., 2000. Consequence of misspecifying assumptions in nonlinear mixed effects models. *Computational Statistics and Data Analysis* 34, 139–164.
- Hausman, J.A., 1978. Specification tests in econometrics. *Econometrica* 46, 1251–1271.
- Heagerty, P.J., Kurland, B.F., 2001. Misspecified maximum likelihood estimates and generalized linear mixed models. *Biometrika* 88, 973–985.
- Heitjan, D.F., Rubin, D.B., 1991. Ignorability and coarse data. *The Annals of Statistics* 19, 2244–2253.
- Horton, N.J., Bechuk, J.D., Jones, C.L., Lipsitz, S.R., Catalano, P.J., Zahner, G.E., Fitzmaurice, G.M., 1999. Goodness-of fit for GEE: an example with mental health service utilization. *Statistics in Medicine* 18, 213–222.
- Hosmer, D.W., Lemeshow, S., 1980. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics, Theory and Method* A9, 1043–1069.
- Huang, X., 2009. Diagnosis of random-effect model misspecification in generalized linear mixed models for binary response. *Biometrics* 65, 361–368.
- Lai, T.L., Shih, M., 2003. Nonparametric estimation in nonlinear mixed effects models. *Biometrika* 90, 1–13.
- Lange, N., Ryan, L., 1989. Assessing normality in random effects models. *Annals of Statistics* 17, 624–642.
- Litière, S., 2007. The impact of a misspecified random-effects distribution on estimation in generalized linear mixed models. Doctoral dissertation. Center for Statistics, Hasselt University, Agoralaan, Belgium.
- Litière, S., Alonso, A., Molenberghs, G., 2007. Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics* 63, 1038–1044.
- Litière, S., Alonso, A., 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* 27, 3125–3144.
- Mardia, K., Kent, J., Bibby, J., 1979. *Multivariate Analysis*. Academic Press.
- McCulloch, C.E., Searle, S.R., Neuhaus, J.M., 2008. *Generalized, Linear, and Mixed Models*. John Wiley & Son, Inc., Hoboken, New Jersey.
- Pinheiro, J.C., Bates, D.M., 1995. Approximation to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 4, 12–35.
- Ritz, C., 2004. Goodness-of-fit tests for mixed models. *Scandinavian Journal of Statistics* 31, 443–458.
- Rotnitzky, A., Wypij, D., 1994. A note on the bias of estimators with missing data. *Biometrics* 50, 419–433.
- Schoenfeld, D., 1980. Chi-square goodness of fit tests for the proportional hazards regression model. *Biometrika* 67, 145–153.
- Tchetgen, E.J., Coull, B.A., 2006. A diagnostic test for the mixing distribution in a generalized linear mixed model. *Biometrika* 93, 1003–1010.
- Tsiatis, A.A., 1980. A note on a goodness-of-fit for the logistic regression. *Biometrika* 67, 250–251.
- Tsiatis, A.A., 2006. *Semiparametric Theory and Missing Data*. Springer, New York.
- Verbeke, G., Lesaffre, E., 1994. Large sample properties of the maximum likelihood estimators in linear mixed models with misspecified random-effects distributions. Report #1996.1. Biostatistical Centre for Clinical Trials, Catholic University of Leuven, Belgium.
- Verbeke, G., Lesaffre, E., 1997. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* 23, 541–556.
- Verbeke, G., Molenberghs, G., 2010. Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling* (submitted for publication).
- Vonesh, E.F., Chinchilli, V.M., Pu, K., 1996. Goodness-of-fit generalized nonlinear mixed-effects models. *Biometrics* 52, 572–587.
- Waagepetersen, R., 2006. A simulation-based goodness-of-fit test for random effects in generalized linear mixed models. *Scandinavian Journal of Statistics* 33, 721–731.
- White, H., 1981. Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* 76, 419–433.
- White, H., 1982. Maximum-likelihood estimation of misspecified models. *Econometrica* 50, 1–25.