Web-based Supplementary Materials for "Diagnosing Misspecification of the Random-Effects Distribution in Mixed Models"

By Reza Drikvandi, Geert Verbeke, and Geert Molenberghs

Web Appendix A

To prove Theorem 1, we first show that the proposed test statistic $T(\hat{\psi})$ is asymptotically equivalent to a quadratic form. According to the large sample theory of ML estimators, if the model is correctly specified, then under general regularity conditions the ML estimate $\hat{\psi}$ of ψ is consistent and asymptotically has a multivariate normal distribution when $N \to \infty$, that is

$$\hat{\psi} \xrightarrow{d} N_L\left(\psi_0, \Sigma\right), \tag{1}$$

in which Σ is the inverse Fisher information matrix of the model parameters (see White, 1982). Note that, because Σ is nonnegative definite, it can be written as $\Sigma = AA'$, where A is the square root of Σ.

By applying a Taylor expansion of $T(\psi)$ around $\hat{\psi}$, we obtain that

$$T(\psi) = T(\hat{\psi}) + (\psi - \hat{\psi})'\nabla(T(\hat{\psi})) + \frac{1}{2}(\psi - \hat{\psi})'\mathrm{H}(T(\hat{\psi}))(\psi - \hat{\psi}) + o_p(1),$$

where ∇ and H are, respectively, the gradient vector and the Hessian matrix of $T(\psi)$. Hence,

$$T(\hat{\psi}) - T(\psi) = (\hat{\psi} - \psi)'\nabla(T(\hat{\psi})) - \frac{1}{2}(\hat{\psi} - \psi)'\mathrm{H}(T(\hat{\psi}))(\hat{\psi} - \psi) + o_p(1).$$

Consequently, for the true parameter vector ψ_0 and using that $T(\psi_0) = 0$ under the null, we would have

$$T(\hat{\psi}) = (\hat{\psi} - \psi_0)'\nabla(T(\hat{\psi})) - \frac{1}{2}(\hat{\psi} - \psi_0)'\mathrm{H}(T(\hat{\psi}))(\hat{\psi} - \psi_0) + o_p(1). \tag{2}$$

On the other hand, since $\nabla(T(\psi_0)) = 0$ under H_0, a Taylor expansion of $\nabla(T(\psi_0))$ around $\hat{\psi}$ gives

$$0 = \nabla(T(\psi_0)) = \nabla(T(\hat{\psi})) + H(T(\hat{\psi}))(\psi_0 - \hat{\psi}) + o_p(1). \tag{3}$$

From (2) and (3), it now follows that

$$T(\hat{\psi}) = \frac{1}{2}(\hat{\psi} - \psi_0)'\mathrm{H}(T(\hat{\psi}))(\hat{\psi} - \psi_0) + o_p(1). \tag{4}$$

We should mention here that $\mathrm{H}(T(\hat{\psi})) \not\to \mathrm{H}(T(\psi_0))$, because $\mathrm{H}(T(\hat{\psi}))$ is a function of Y_i irrespective of $\hat{\psi}$.

Since, under H_0, for each b

$$\hat{\Delta}(\hat{G}, b) - 1 = o_p(1)$$

and

$$\frac{\partial}{\partial \psi_l}\Delta\left(G, b\right)\Big|_{\psi=\hat{\psi}} - \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{\partial}{\partial \psi_{0l}}\frac{f_i(Y_i|b)}{f_i(Y_i|G)}\right] = o_p(1),$$

it is straightforward to show that

$$\frac{\partial^2}{\partial \psi_l \partial \psi_{l'}} T(\psi)\big|_{\psi=\hat{\psi}} \stackrel{P}{\to} \int_{R^q} 2\left\{ \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left[\frac{\partial}{\partial \psi_{0l}} \frac{f_i(Y_i|b)}{f_i(Y_i|G)} \right] \right\} \left\{ \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left[\frac{\partial}{\partial \psi_{0l'}} \frac{f_i(Y_i|b)}{f_i(Y_i|G)} \right] \right\} dG(b).$$

Thus,

$$\frac{1}{2}\mathrm{H}(T(\hat{\psi})) \stackrel{P}{\to} Q(\psi_0), \tag{5}$$

in which $Q(\psi_0)$ is defined as in the statement of Theorem 1.

Now, by using (1) and (5) and applying Slutsky's theorem, we obtain from (4) that

$$T(\hat{\psi}) = (AZ)'\, Q(\psi_0)\, (AZ) + o_p(1),$$

where Z is the standard multivariate normal distribution. But

$$(AZ)'\, Q(\psi_0)\, (AZ) = Z'A'Q(\psi_0)AZ.$$

Since $A'Q(\psi_0)A$ is a symmetric matrix, by using the spectral decomposition we can write it as $PD(\lambda_i)P'$, where P and $D(\lambda_i)$ are, respectively, the orthogonal matrix of eigenvectors and the diagonal matrix of eigenvalues of $A'Q(\psi_0)A$. Hence,

$$Z'A'Q(\psi_0)AZ = Z'PD(\lambda_i)P'Z = (P'Z)'D(\lambda_i)P'Z.$$

$P'Z$ is also distributed as the standard multivariate normal distribution. We thus have

$$(P'Z)'D(\lambda_i)P'Z \sim Z'D(\lambda_i)Z \sim \sum_{i=1}^{r} \lambda_i \chi_i^2,$$

where χ_i^2 $(i = 1, ..., r)$ are independent χ_1^2 random variables, and this completes the proof.

Web Appendix B

Critical values of the proposed test statistic $T(\hat{\psi})$ can be computed analytically as follows. Under the conditions of Theorem 1, we have

$$P(T(\hat{\psi}) > t) = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \frac{\sin \omega(u)}{uh(u)} du, \qquad (6)$$

where $\omega(u) = (\sum_{j=1}^{r} tg^{-1}\lambda_j u - tu)/2$ and $h(u) = \prod_{j=1}^{r} \left(1 + \lambda_j^2 u^2\right)^{1/4}$.

The function $uh(u)$ in the denominator increases monotonically towards $+\infty$, therefore the integration in (6) can be carried over a finite range $0 \leq u \leq U$. It has been shown that the error of truncation is bounded by $\frac{2}{\pi r} U^{2/r} \prod_{j=1}^{r} \lambda_j^{-1/2}$ (Imhof, 1961, p. 423). In our work, we choose U such that the error of truncation is less than $\epsilon = 0.0001$. Moreover, since the integrand in (6) is a periodic function, we propose to calculate the definite integral by using the trapezoidal rule with steps of length 1/10 (see also Imhof, 1961).

Web Appendix C

Our simulation study in Section 7 of the paper concerned misspecification regarding the form of the random-effects distribution. Heagerty and Kurland (2001) demonstrated other forms of misspecification related to the random-effects part, such as ignoring a random effect, group-specific variances, and autoregressive random effects, that could impact inference on fixed-effects parameters. Here, we conducted some further simulations to examine whether our diagnostic test has good power to detect such types of misspecification. Specifically, we conducted simulations to detect misspecifications regarding autoregressive random effects as well as ignoring a random effect. We again considered the logistic mixed model (7). First, for the case of ignoring a random effect we additionally generated a random slope associated with covariate x_{ij} with zero mean and variance 1 to be ignored incorrectly when fitting the model. The simulation results, presented in Table 1, show that our test is considerably more powerful than the test of Tchetgen and Coull (2006) to detect misspecification regarding the ignorance of this random slope, while the test of Alonso et al. (2008) behaves very well in this case and outperforms our asymptotic test when the sample size is not large, but when the sample size is large ($N > 300$) our test performs almost the same as their test. Next, for the case of autoregressive random effects we generated random effects b_{ij} such that $\text{cov}(b_{ij}, b_{ik}) = \sigma^2 \rho^{|j-k|}$, where the correlation parameter ρ was set to 0.5. The simulation results, presented in Table 1, indicate that our test is not powerful enough to detect misspecification regarding autoregressive random effects. The test of Alonso et al. (2008) performs better, however its power is not very large, while again the test of Tchetgen and Coull (2006) does not perform well in this case. Note that since the test of Tchetgen and Coull (2006) relies on the existence of at least one time-varying covariate in the model, we have found that their test performs well for situations where there are several time-varying covariates in the model because then, according to their methodology, the difference between the conditional and marginal estimates could be larger and hence their test would be more powerful.

Table 1: Power of our diagnostic test, denoted by T, the test based on the adjusted test statistic, denoted by T^*, the determinant-trace test of Alonso et al. (2008), denoted by δ_{dt}, and the test of Tchetgen and Coull (2006), denoted by D, to detect two types of misspecification: ignoring a random effect and autoregressive random effects. Note that in each case a normal distribution was assumed to fit the model. Also, RE is just abbreviation of random effect.

		$N = 100$		$N = 200$		$N = 300$		$N = 500$		$N = 1000$	
Type of misspecification		$n = 10$	$n = 15$	$n = 10$	$n = 15$	$n = 10$	$n = 15$	$n = 10$	$n = 15$	$n = 10$	$n = 15$
Ignoring a RE	T	0.018	0.043	0.091	0.273	0.391	0.504	0.690	0.713	0.878	0.919
	T^*	0.048	0.130	0.175	0.362	0.465	0.611	0.706	0.764	0.916	0.944
	δ_{dt}	0.116	0.152	0.239	0.491	0.434	0.609	0.692	0.739	0.892	0.920
	D	0.087	0.091	0.127	0.143	0.130	0.201	0.265	0.294	0.315	0.397
Autoregressive RE	T	0.009	0.012	0.067	0.081	0.095	0.116	0.148	0.175	0.203	0.247
	T^*	0.010	0.043	0.095	0.104	0.123	0.178	0.186	0.209	0.244	0.281
	δ_{dt}	0.086	0.098	0.111	0.125	0.143	0.210	0.214	0.267	0.292	0.326
	D	0.010	0.014	0.053	0.075	0.087	0.099	0.103	0.118	0.162	0.180

In general, our diagnostic test performs very well when the interest is to detect misspecification of the random-effects distribution, and moreover it is able to detect other forms of misspecification such as ignoring some random effect from the model, though it did not show a good power in detecting autoregressive random effects but it was observed that both the test of Tchetgen and Coull (2006) and the test of Alonso et al. (2008) did not perform well in this case. One possible reason is that the autoregressive random effects were generated from a normal distribution and since the model was fitted assuming normal random effects, none of the tests was able to detect the incorrect covariance structure of the model. A test for covariance structure should be more powerful in this case.

References

Alonso, A., Litière, S. and Molenberghs, G. (2008), 'A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models', *Computational*

Statistics and Data Analysis **52**, 4474–4486.

Heagerty, P. J. and Kurland, B. F. (2001), 'Misspecified maximum likelihood estimates and generalised linear mixed models', *Biometrika* **88**, 973–985.

Imhof, J. P. (1961), 'Computing the distribution of quadratic forms in normal variables', *Biometrika* **48**, 419–426.

Tchetgen, E. J. and Coull, B. A. (2006), 'A diagnostic test for the mixing distribution in a generalised linear mixed model', *Biometrika* **93**, 1003–1010.

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