Biometrika (2023), **105**, 1, *pp.* 1–8
Printed in Great Britain Advance Access publication on 31 December 2023

Second term improvement to generalised linear mixed models asymptotics

BY LUCA MAESTRINI
Research School of Finance, Actuarial Studies and Statistics, The Australian National University, Canberra 2601, Australia
luca.maestrini@anu.edu.au

AISHWARYA BHASKARAN AND MATT P. WAND
School of Mathematical and Physical Sciences, University of Technology Sydney, P.O. Box 123, Broadway 2007, Australia
aishwarya.bhaskaran@mq.edu.au, matt.wand@uts.edu.au

SUMMARY

A recent article on generalised linear mixed model asymptotics, Jiang *et al.* (2022), derived the rates of convergence for the asymptotic variances of maximum likelihood estimators. If m denotes the number of groups and n is the average within-group sample size then the asymptotic variances have orders m^{-1} and $(mn)^{-1}$, depending on the parameter. We extend this theory to provide explicit forms of the $(mn)^{-1}$ second terms of the asymptotically harder-to-estimate parameters. Improved accuracy of statistical inference and planning are consequences of our theory.

Some key words: Longitudinal data analysis, Maximum likelihood estimation, Sample size calculations.

1. INTRODUCTION

Generalised linear mixed models are a vehicle for regression analysis of grouped data with non-Gaussian responses such as counts and categorical labels. Until recently, the precise asymptotic behaviours of the conditional maximum likelihood estimators were not known for these models. Jiang *et al.* (2022) derived leading term asymptotic variances and showed that they have orders m^{-1} and $(mn)^{-1}$, depending on the parameter, where m is the number of groups and n is the average within-group sample size. The main contribution of this article is to extend the asymptotic variance and covariance approximations to terms in $(mn)^{-1}$ for *all* parameters. This constitutes *second term improvement* to generalized linear mixed model asymptotics. The potential statistical payoffs are improved accuracy of confidential intervals, hypothesis tests, sample size calculations and optimal design.

The essence of generalized linear mixed models is the extension of general linear models via the addition of random effects that allow for the handling of correlations arising from repeated measures. There are numerous types of random effect structures. The most common is the two-level nested structure, corresponding to repeated measures within each of m distinct groups. This version of generalised linear mixed models, with frequentist inference via maximum likelihood and its quasi-likelihood extension, is our focus here.

Suppose that a fixed effects parameter in a two-level generalised linear mixed model is accompanied by a random effect. Jiang *et al.* (2022) showed that the variance of its maximum likelihood estimator, conditional on the predictor data, is asymptotic to $C_1 m^{-1}$ for some deterministic constant C_1 that depends on the true model parameter values. The crux of this article is to extend the asymptotic variance approximation to $C_1 m^{-1} + C_2 (mn)^{-1}$ for an additional deterministic constant C_2. We derive the explicit form of C_2 for two-level nested generalised linear mixed models for both maximum likelihood and maximum quasi-likelihood situations. Even though, in general, C_2 does not have a succinct form it is still usable in that operations such as studentisation are straightforward and result in improvements in statistical utility.

For two-level nested mixed models, $(mn)^{-1}$ is the best possible rate of convergence for the asymptotic variance of the estimator of a model parameter. Such a rate is achieved by maximum likelihood estimators of fixed effects parameters unaccompanied by random effects and dispersion parameters (e.g. Bhaskaran & Wand, 2023). The current article closes the problem of obtaining the precise asymptotic forms of the variances, up to terms in $(mn)^{-1}$, for estimation of *all* model parameters. The achievement of this theoretical outcome has required very many algebraic steps and order of magnitude verifications beyond those given in Jiang *et al.* (2022). For example, three-dimensional arrays and their combination with regular matrices play a central role. We introduce a new type of array multiplication that streamlines the required manipulations.

§ 2 describes the model under consideration and corresponding maximum likelihood estimators. Our second term improvement results are presented in § 3. § 4 describes statistical utility due to the new asymptotic results. A supplement to this article contains derivational and various auxiliary details.

2. MODEL DESCRIPTION AND MAXIMUM LIKELIHOOD ESTIMATION

Consider the class of two-parameter exponential family of density, or probability mass, functions with generic form

$$p(y;\eta,\phi) = \exp[\{y\eta - b(\eta) + c(y)\}/\phi + d(y,\phi)]h(y) \tag{1}$$

where η is the *natural parameter* and $\phi > 0$ is the *dispersion parameter*. Examples include the Gaussian density for which $b(x) = \frac{1}{2}x^2$, $c(x) = -\frac{1}{2}x^2$, $d(x_1, x_2) = -\frac{1}{2}\log(2\pi x_2)$ and $h(x) = I(x \in \mathbb{R})$ and the Gamma density function for which $b(x) = -\log(-x)$, $c(x) = \log(x)$, $d(x_1, x_2) = -\log(x_1) - \log(x_2)/x_2 - \log\Gamma(1/x_2)$ and $h(x) = I(x > 0)$. Here $I(\mathcal{P}) = 1$ if the condition $\mathcal{P}$ is true and $I(\mathcal{P}) = 0$ if $\mathcal{P}$ is false. The Binomial and Poisson probability mass functions are also special cases of (1) but with ϕ fixed at 1. When (1) is used in regression contexts a common modelling extension for count and proportion responses, usually to account for overdispersion, is to remove the $\phi = 1$ restriction and replace it with $\phi > 0$. In these circumstances

$$\{y\eta - b(\eta) + c(y)\}/\phi + d(y,\phi) \tag{2}$$

is labelled a *quasi-likelihood function* since it is not the logarithm of a probability mass function for $\phi \neq 1$. We use the more general quasi-likelihood terminology for the remainder of this article.

Consider, for observations of the random pairs (X_{ij}, Y_{ij}), $1 \le i \le m$, $1 \le j \le n_i$, generalised linear mixed models of the form,

$$Y_{ij}|X_{ij}, U_i \text{ independent having quasi-likelihood function (2) with natural parameter} \\ \left(\beta^0 + \begin{bmatrix} U_i \\ 0 \end{bmatrix}\right)^T X_{ij} \text{ such that the } U_i \text{ are independent } N(0, \Sigma^0) \text{ random vectors.} \tag{3}$$

The X_{ij} are $d_{\rm F} \times 1$ random vectors corresponding to predictors. The U_i are $d_{\rm R} \times 1$ unobserved random effects vectors, where $d_{\rm R} \le d_{\rm F}$. Under this set-up the first $d_{\rm R}$ entries of the X_{ij} are partnered by a random effect. The remaining entries correspond to predictors that have a fixed effect only. We assume that the X_{ij} and U_i, for $1 \le i \le m$ and $1 \le j \le n_i$, are totally independent, with the X_{ij} each having the same distribution as the $d_{\rm F} \times 1$ random vector X and the U_i each having the same distribution as the $d_{\rm R} \times 1$ random vector U.

For any β ($d_{\rm F} \times 1$) and Σ ($d_{\rm R} \times d_{\rm R}$) that is symmetric and positive definite and conditional on the X_{ij} data, the quasi-likelihood is

$$\ell(\beta, \Sigma) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} [\{Y_{ij}(\beta^T X_{ij} + c(Y_{ij})\}/\phi + d(Y_{ij}, \phi)] - \frac{m}{2} \log |2\pi\Sigma| \\ + \sum_{i=1}^{m} \log \int_{\mathbb{R}^{d_{\rm R}}} \exp\left[\frac{1}{\phi} \sum_{j=1}^{n_i} \left\{ Y_{ij} \begin{bmatrix} u \\ 0 \end{bmatrix}^T X_{ij} - b\left(\left(\beta + \begin{bmatrix} u \\ 0 \end{bmatrix}\right)^T X_{ij}\right)\right\} - \tfrac{1}{2} u^T \Sigma^{-1} u\right] du.$$

The maximum quasi-likelihood estimator of (β^0, Σ^0) is $(\widehat{\beta}, \widehat{\Sigma}) = \text{argmax}_{\beta,\Sigma}\, \ell(\beta, \Sigma)$. In practice computation of $(\widehat{\beta}, \widehat{\Sigma})$ can be challenging due to intractable $d_{\rm R}$-dimensional integrals, although ongoing advances tend to alleviate this problem. We ignore this aspect here and study the theoretical properties of the exact maximum quasi-likelihood estimator rather than approximations to them.

Suppose that $d_{\rm F} > d_{\rm R}$ and consider the partition $\beta = [\beta_{\rm A}^T \ \beta_{\rm B}^T]^T$ of the fixed effects parameter vector, where $\beta_{\rm A}$ is $d_{\rm R} \times 1$ and $\beta_{\rm B}$ is $(d_{\rm F} - d_{\rm R}) \times 1$. The $d_{\rm F} = d_{\rm R}$ boundary case is such that $\beta_{\rm B}$ is null. Also, let $\mathcal{X} \equiv \{X_{ij} : 1 \le i \le m,\ 1 \le j \le n_i\}$. Theorem 1 of Jiang *et al.* (2022) implies that, under some mild conditions, the covariance matrices of $\widehat{\beta}_{\rm A}$, $\widehat{\beta}_{\rm B}$ and $\text{vech}(\widehat{\Sigma})$ have leading term behaviour given by

$$\text{Cov}\big(\widehat{\beta}_{\rm A}|\mathcal{X}\big) = \frac{\Sigma^0\{1 + o_p(1)\}}{m}, \ \ \text{Cov}\big(\widehat{\beta}_{\rm B}|\mathcal{X}\big) = \frac{\phi\Lambda_{\beta_{\rm B}}\{1 + o_p(1)\}}{mn}, \tag{4}$$

where $n \equiv \frac{1}{m}\sum_{i=1}^{m} n_i$, and

$$\text{Cov}\big(\text{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D_{d_{\rm R}}^{+}(\Sigma^0 \otimes \Sigma^0)D_{d_{\rm R}}^{+T}\{1 + o_p(1)\}}{m}. \tag{5}$$

Here $\Lambda_{\beta_{\rm B}}$ is a $(d_{\rm F} - d_{\rm R}) \times (d_{\rm F} - d_{\rm R})$ matrix that depends on β and the (X, U) distribution, $D_{d_{\rm R}}$ is the matrix of zeroes and ones such that $D_{d_{\rm R}}\text{vech}(A) = \text{vec}(A)$ for all $d_{\rm R} \times d_{\rm R}$ symmetric matrices A and $D_{d_{\rm R}}^{+} = (D_{d_{\rm R}}^T D_{d_{\rm R}})^{-1} D_{d_{\rm R}}^T$ is the Moore-Penrose inverse of $D_{d_{\rm R}}$. The theory of Jiang *et al.* (2022) also indicates a degree of asymptotic orthogonality between $\beta_{\rm A}$ and $\beta_{\rm B}$ in that $E\big\{(\widehat{\beta}_{\rm A} - \beta_{\rm A}^0)(\widehat{\beta}_{\rm B} - \beta_{\rm B}^0)^T|\mathcal{X}\big\}$ has $O_p\{(mn)^{-1}\}$ entries, which implies that the correlations between the entries of $\widehat{\beta}_{\rm A}$ and $\widehat{\beta}_{\rm B}$ are asymptotically negligible. For Gaussian responses, Lyu & Welsh (2022) considered an extension of (3) for which some entries of X_{ij} are constrained to be constant across all n_i measurements within the ith group. For such constant-within-group predictors they showed that the asymptotic variances of the corresponding fixed effects parameters are of order m^{-1} rather than $(mn)^{-1}$. This type of extension is not considered here, but is worthy of future consideration.

The leading term approximations of the variability in $\widehat{\beta}_{\rm A}$ and $\text{vech}(\widehat{\Sigma})$, given by (4) and (5), are somewhat crude. Unlike the asymptotic covariance of $\widehat{\beta}_{\rm B}$, they do not show the effect of the average within-group sample size n. In the next section we investigate their second term improvements.

3. TWO-TERM ASYMPTOTIC COVARIANCE RESULTS

We define the two-term asymptotic covariance matrix problem to be the determination of the *unique deterministic* matrices M_β and M_Σ such that

$$\text{Cov}\big(\widehat{\beta}|\mathcal{X}\big) = \frac{1}{m}\begin{bmatrix}\Sigma^0 & O\\ O & O\end{bmatrix} + \frac{M_\beta\{1+o_p(1)\}}{mn} \quad \text{and}$$

$$\text{Cov}\big(\text{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D_{d_{\rm R}}^{+}(\Sigma^0\otimes\Sigma^0)D_{d_{\rm R}}^{+T}}{m} + \frac{M_\Sigma\{1+o_p(1)\}}{mn}$$

under reasonably mild conditions.

An example for which a solution to the two-term asymptotic covariance problem can be expressed relatively simply is the $d_{\rm F}=2$, $d_{\rm R}=1$ Poisson quasi-likelihood special case of (3), with parameters

$$\beta = (\beta_0,\beta_1) \quad \text{and} \quad \Sigma = \sigma^2 \quad \text{and predictor variable} \quad X = [1\ X]^T$$

for a scalar random variable X. Define

$$a_1\big(\beta_0,\beta_1,\sigma^2\big) \equiv e^{\beta_0+\sigma^2/2}\big[E(X^2e^{\beta_1 X})E(e^{\beta_1 X}) - \{E(Xe^{\beta_1 X})\}^2\big]$$

and

$$a_2(\beta_1,\sigma^2) \equiv \frac{e^{\sigma^2}E\big(X^2e^{\beta_1X}\big)E\big(e^{\beta_1X}\big) + \big(1-e^{\sigma^2}\big)E\{\big(Xe^{\beta_1X}\big)\}^2}{E\big(e^{\beta_1X}\big)}.$$

Then the two-term covariance matrix of $(\widehat{\beta}_0,\widehat{\beta}_1)$ is

$$\text{Cov}\left(\begin{bmatrix}\widehat{\beta}_0\\ \widehat{\beta}_1\end{bmatrix}\Bigg|\mathcal{X}\right) = \frac{1}{m}\begin{bmatrix}(\sigma^2)^0 & 0\\ 0 & 0\end{bmatrix} + \frac{\phi\{1+o_p(1)\}}{a_1\big(\beta_0^0,\beta_1^0,(\sigma^2)^0\big)\,mn}\begin{bmatrix}a_2\big(\beta_1^0,(\sigma^2)^0\big) & -E\big(Xe^{\beta_1^0X}\big)\\ -E\big(Xe^{\beta_1^0X}\big) & E\big(e^{\beta_1^0X}\big)\end{bmatrix}.$$

Studentisation of the two-term asymptotic covariance matrix for obtaining confidence intervals and Wald hypothesis tests is straightforward. For example, $E(X^2e^{\beta_1^0X})$ can be replaced by the estimator $(mn)^{-1}\sum_{i=1}^m\sum_{j=1}^{n_i}X_{ij}^2e^{\widehat{\beta}_1X_{ij}}$.

The remainder of this section is concerned with the *theoretical* problem of obtaining the forms of M_β and M_Σ for model (3) in general. The achievement of this goal has turned out to be quite challenging. The score asymptotic approximation approach used in Jiang *et al.* (2022) requires higher numbers of terms to obtain valid two-term covariance matrix approximations. Some of these terms can only be expressed using three-dimensional arrays rather than with matrices. Succinct statement of M_β and M_Σ is only possible with well-designed nested function notation. A novel notation for multiplicative combining of three-dimensional arrays with compatible matrices is also beneficial. The next subsection focusses on these notational aspects.

3.1. *Notation for the Main Result*

Let $\mathcal{A}$ be a $d_1 \times d_2 \times d_3$ array and M be a $d_1 \times d_2$ matrix. Then we let

$$\mathcal{A}\bigstar M \quad \text{denote the } d_3 \times 1 \text{ vector with } t\text{th entry given by} \quad \sum_{r=1}^{d_1}\sum_{s=1}^{d_2}(\mathcal{A})_{rst}(M)_{rs}. \tag{6}$$

Next, for $U \sim N(0, \Sigma^0)$, define

$$\Omega_{\text{AA}}(U) \equiv E\Big\{b''\big((\beta_{\text{A}}^0 + U)^T X_{\text{A}} + (\beta_{\text{B}}^0)^T X_{\text{B}}\big)X_{\text{A}}X_{\text{A}}^T|U\Big\},$$

$$\Omega_{\text{AB}}(U) \equiv E\Big\{b''\big((\beta_{\text{A}}^0 + U)^T X_{\text{A}} + (\beta_{\text{B}}^0)^T X_{\text{B}}\big)X_{\text{A}}X_{\text{B}}^T|U\Big\}$$

$$\text{and} \quad \Omega_{\text{BB}}(U) \equiv E\Big\{b''\big((\beta_{\text{A}}^0 + U)^T X_{\text{A}} + (\beta_{\text{B}}^0)^T X_{\text{B}}\big)X_{\text{B}}X_{\text{B}}^T|U\Big\}.$$

Also let $\Omega'_{\text{AAA}}(U)$ be the $d_{\text{R}} \times d_{\text{R}} \times d_{\text{R}}$ array with (r, s, t) entry equal to

$$E\Big\{b'''\big((\beta_{\text{A}}^0 + U)^T X_{\text{A}} + (\beta_{\text{B}}^0)^T X_{\text{B}}\big)(X_{\text{A}})_r(X_{\text{A}})_s(X_{\text{A}})_t\big|U\Big\}.$$

and $\Omega'_{\text{AAB}}(U)$ be the $d_{\text{R}} \times d_{\text{R}} \times (d_{\text{F}} - d_{\text{R}})$ array with (r, s, t) entry equal to

$$E\Big\{b'''\big((\beta_{\text{A}}^0 + U)^T X_{\text{A}} + (\beta_{\text{B}}^0)^T X_{\text{B}}\big)(X_{\text{A}})_r(X_{\text{A}})_s(X_{\text{B}})_t\big|U\Big\}.$$

Define the random vectors:

$$\psi_1(U) \equiv \text{vech}(\Sigma - UU^T), \;\; \psi_2(U) \equiv \Omega'_{\text{AAA}}(U)\bigstar\Omega_{\text{AA}}(U)^{-1}, \;\; \psi_3(U) \equiv \Omega'_{\text{AAB}}(U)\bigstar\Omega_{\text{AA}}(U)^{-1}$$

$$\text{and} \;\; \psi_4(U) \equiv D_{d_{\text{R}}}^+ \text{vec}\Big(\Omega_{\text{AA}}(U)^{-1}\Sigma^{-1}\big\{\Sigma - UU^T - \Sigma\psi_2(U)U^T\big\}\Big).$$

Then define the random matrices:

$$\Psi_5(U) \equiv \Omega_{\text{AA}}(U)^{-1}\Omega_{\text{AB}}(U), \quad \Psi_6(U) \equiv \Omega_{\text{BB}}(U) - \Psi_5(U)^T\Omega_{\text{AB}}(U),$$

$$\Psi_7(U) \equiv UU^T\Sigma^{-1}\Omega_{\text{AA}}(U)^{-1}, \quad \Psi_8(U) \equiv D_{d_{\text{R}}}^+\big[(UU^T) \otimes \{\Omega_{\text{AA}}(U)^{-1}\}\big]D_{d_{\text{R}}}^{+T}$$

$$\text{and} \quad \Psi_9(U) \equiv \psi_1(U)\psi_4(U)^T + \psi_4(U)\psi_1(U)^T.$$

Lastly, define the expectation matrices:

$$\Lambda_{\text{AA}} \equiv E\Big\{\Psi_7(U) + \Psi_7(U)^T - \Omega_{\text{AA}}(U)^{-1} + \Omega_{\text{AA}}(U)^{-1}\psi_2(U)U^T + U\psi_2(U)^T\Omega_{\text{AA}}(U)^{-1}\Big\},$$

$$\Lambda_{\text{AB}} \equiv E\Big\{UU^T\Sigma^{-1}\Psi_5(U) + U\psi_2(U)^T\Psi_5(U) - U\psi_3(U)^T\Big\} \quad \text{and}$$

$$\Delta \equiv E\Big(\Big[\Psi_5(U)^T\big\{\Sigma^{-1}U + \psi_2(U)\big\} - \psi_3(U)\Big]\psi_1(U)^T\Big).$$

3.2. *Assumptions for the Main Result*

The main result depends on the following sample size asymptotic assumptions: the number of groups m diverges to ∞; the within-group sample sizes n_i diverge to ∞ in such a way that $n_i/n \to C_i$ for constants $0 < C_i < \infty$, $1 \le i \le m$; the ratio n/m converges to zero. The last of these conditions is in keeping with the number of groups being large compared with the within-group sample sizes, as often arises in practice. For our asymptotics it ensures that, for the harder-to-estimate parameters, the asymptotic variances of the maximum likelihood estimators have leading terms of the form $C_1 m^{-1} + C_2(mn)^{-1}$. In addition, it ensures that the Fisher information is sufficiently dominant for obtaining asymptotic variances.

We also assume that the (X, U) joint distribution is such that all required convergence in probability limits that appear in the deterministic order $(mn)^{-1}$ terms are justified. The determination of sufficient conditions on the (X, U) distribution that guarantee the validity of the main result is a tall order, involving the determination of at least eighteen additional moment-type conditions for results similar to Lemma A1 of Jiang *et al.* (2022), and beyond the scope of this article.

3.3. *Statement of the Main Result*

Using the notation presented in § 3.1 and under the assumptions described in § 3.2, and assuming $d_{\text{F}} > d_{\text{R}}$ we have

$$\text{Cov}\big(\widehat{\beta}|\mathcal{X}\big) = \frac{1}{m}\begin{bmatrix}\Sigma^0 & O\\ O & O\end{bmatrix} + \frac{\phi}{mn}\begin{bmatrix}\Lambda_{\text{AA}}^{-1} & \Lambda_{\text{AA}}^{-1}\Lambda_{\text{AB}}\\ \Lambda_{\text{AB}}^T\Lambda_{\text{AA}}^{-1} & \Lambda_{\text{AB}}^T\Lambda_{\text{AA}}^{-1}\Lambda_{\text{AB}} + E\big\{\Psi_6(U)\big\}\end{bmatrix}^{-1}\{1+o_p(1)\}$$

$$\text{and Cov}\big(\text{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D_{d_{\text{R}}}^{+}(\Sigma^0\otimes\Sigma^0)D_{d_{\text{R}}}^{+T}}{m} + \frac{\phi}{mn}\Big(2E\big\{\Psi_9(U) - 2\Psi_8(U)\big\} + \Delta^T\big[E\{\Psi_6(U)\}\big]^{-1}\Delta\Big)\{1+o_p(1)\}. \tag{7}$$

For the $d_{\text{F}} = d_{\text{R}}$ boundary case the first term of $\text{Cov}\big(\widehat{\beta}|\mathcal{X}\big)$ is simply $\frac{1}{m}\Sigma^0$. A supplement to this article contains a full derivation of (7).

In the Gaussian response special case we have $b''(x) = 1$ and $b'''(x) = 0$ and the main result reduces to the following succinct form:

$$\text{Cov}\big(\widehat{\beta}|\mathcal{X}\big) = \frac{1}{m}\begin{bmatrix}\Sigma^0 & O\\ O & O\end{bmatrix} + \frac{\phi\big\{E(XX^T)\big\}^{-1}\{1+o_p(1)\}}{mn} \quad \text{and}$$

$$\text{Cov}\big(\text{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D_{d_{\text{R}}}^{+}(\Sigma^0\otimes\Sigma^0)D_{d_{\text{R}}}^{+T}}{m} + \frac{4\phi D_{d_{\text{R}}}^{+}\big[\Sigma^0\otimes\{E(X_{\text{A}}X_{\text{A}}^T)\}^{-1}\big]D_{d_{\text{R}}}^{+T}\{1+o_p(1)\}}{mn}.$$

4. UTILITY OF THE SECOND TERM IMPROVEMENTS

The second term improvements of (7) have ready and straightforward applications to confidence intervals, Wald hypothesis tests and sample size calculations. Optimal design is another possible utility, but would require second term improvements of the type of theory given in § 5 of Jiang *et al.* (2022).

We conducted a simulation exercise aimed at understanding potential practical impacts of second term improvements to generalized linear mixed model asymptotics. Data sets were generated from the $d_{\text{F}} = 5$ and $d_{\text{R}} = 2$ logistic mixed model

$$Y_{ij}|X_{1ij}, X_{2ij}, X_{3ij}, X_{4ij}, U_i \text{ independently distributed as}$$
$$\text{Bernoulli}\Big(1/\big(1+\exp[-\{\beta_0^0 + U_{0i} + (\beta_1^0 + U_{1i})X_{1ij} + \beta_2^0 X_{2ij} + \beta_3^0 X_{3ij} + \beta_4^0 X_{4ij}\}]\big)\Big),$$
$$\text{where the } \begin{bmatrix}U_{0i}\\ U_{1i}\end{bmatrix} \text{ are independent } N(0,\Sigma^0) \text{ random vectors, } 1\le i\le m,\ 1\le j\le n.$$

The 'true' parameter values were set to

$$\left(\beta_0^0,\beta_1^0,\beta_2^0,\beta_3^0,\beta_4^0\right)=(0.35,0.96,-0.47,1.06,-1.31)\quad\text{and}\quad\Sigma^0=\begin{bmatrix}0.56 & -0.34\\ -0.34 & 0.89\end{bmatrix}$$

and the predictor data were generated from independent Uniform distributions on the unit interval. The simulation design is such that the asymptotic variance of $\widehat{\beta}_1$, corresponding to the fixed effect of the X_1 predictor, benefits from second term improvement. The true β^0 vector was chosen so that there was a variety of strengths of predictor fixed effects. We selected the Σ^0 matrix to ensure that there was a significant amount of heterogeneity in the random intercepts and slopes. In our reporting of simulation results we use the following standard deviation and correlation parameterisation: $\sigma_1^0\equiv(\Sigma_{11}^0)^{1/2}$, $\sigma_2^0\equiv(\Sigma_{22}^0)^{1/2}$ and $\rho^0\equiv\Sigma_{12}^0/(\sigma_1^0\sigma_2^0)$. To assess potential large sample improvements afforded by the two-term asymptotic covariance expressions at (7) we varied m over the set $\{100,150,\ldots,500\}$ and fixed n at $m/10$. For each (m,n) pair we then simulated 500 replications and obtained approximate 95% confidence intervals for all model parameters according to the approach described in § 4 of Jiang *et al.* (2022) and the second term improvements arising from (7). The confidence intervals for σ_1^0 and σ_2^0 involved use of asymptotic normality results for logarithms of these parameters, followed by exponentiation. Similar remarks apply to ρ^0 but with use of $\tanh^{-1}$ and $\tanh$ functions. The requisite bivariate integrals were obtained using the function `hcubature()` within the R language package cubature (Balasubramanian *et al.*, 2023). The point estimates, which were obtained via the R language package glmmTMB (Brooks *et al.*, 2023), use Laplace's method to approximate bivariate integrals.

Note that the confidence intervals for β_0^0, β_1^0 and the entries of Σ^0 differ according to the two asymptotic theory approaches since the estimators of these parameters have order m^{-1} asymptotic variances. The confidence intervals for β_2^0, β_3^0 and β_4^0 are unaffected by the second term asymptotic improvements since their estimators have order $(mn)^{-1}$ asymptotic variances.

Figure 1 compares the empirical coverages of confidence intervals with advertised levels of 95% for the one-term asymptotic variances of Jiang *et al.* (2022) and the two-term asymptotic variances that arise from (7). In Figure 1 we only consider the parameters that are affected by second term improvement. The empirical coverages for the other parameters are provided in the supplement. For comparison with existing software products, the empirical coverages for the glmmTMB confidence intervals are also shown in Figure 1. For β_0^0, β_1^0 and σ_1^0 there is close correspondence between the two-term and glmmTMB confidence intervals. For σ_2^0 and ρ^0 and lower values of m, the two-term confidence intervals are prone to some under-coverage whilst glmmTMB has empirical coverages above the advertised level.

It is clear from Figure 1 that our second term improvements lead to much better coverages for lower sample size situations. On the other hand, one-term confidence intervals are trivial to compute whilst the two-term versions require considerable computing involving numerical integration.

Simulation results such as those summarised by Figure 1 provide an appreciation for the practical trade-offs arising from precise asymptotics for generalised linear mixed models.

ACKNOWLEDGEMENTS

We are grateful to Alessandra Salvan and Nicola Sartori for advice related to this research. This research was supported by the Australian Research Council Discovery Project DP230101179.

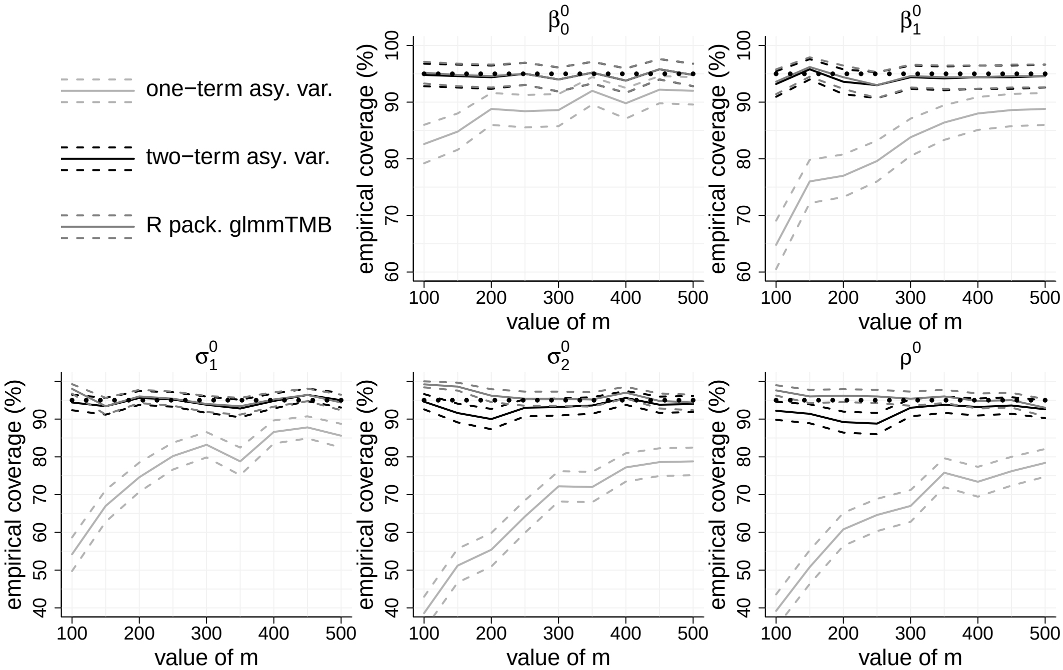

Fig. 1. *Empirical coverage of confidence intervals from the simulation exercise described in the text. Each panel corresponds to a model parameter that is impacted by second term asymptotic improvements. The advertised coverage level is fixed at 95% and is indicated by a horizontal dotted line in each panel. The solid curves show, dependent on the number of groups m, the empirical coverage levels for confidence intervals based on each of the three approaches. The dashed curves correspond to plus and minus two standard errors of the sample proportions. The within-group sample size, n, is fixed at $m/10$.*

Supplementary material

Supplementary material available at *Biometrika* online contains derivational details and additional simulation results.

References

BALASUBRAMANIAN, N., JOHNSON, S.G., HAHN, T., BOUVIER, A. & KIÊU, K. (2023). cubature 2.0.4.6: Adaptive multivariate integration over hypercubes. R package.

BHASKARAN, A. AND WAND, M.P. (2023). Dispersion parameter extension of precise generalized linear mixed model asymptotics. *Statist. Probab. Lett.*, **193**, Article 109691.

BROOKS, M., BOLKER, B., KRISTENSEN, K., MAECHLER, M., MAGNUSSON, A., SKAUG, H. NIELSEN, A., BERG, C., VAN BENTHAM, K. (2023). glmmTMB 1.1.7: Generalized linear mixed models using Template Model Builder. R package.

JIANG, J., WAND, M.P. & BHASKARAN, A. (2022). Usable and precise asymptotics for generalized linear mixed model analysis and design. *J. R. Statist. Soc., Ser. B*, **84**, 55–82.

LYU, Z. & WELSH, A.H. (2022). Increasing cluster size asymptotics for nested error regression models. *J. Statist. Plan. Inf.*, **217**, 52–68.

[*Received on* 31 *March* 2023. *Editorial decision on* 31 *December* 9999]