**Noname manuscript No.**
(will be inserted by the editor)

# Simplified Partially Observed Quasi-Information Matrix

**Thuan Nguyen · Jiming Jiang**

Received: date / Accepted: date

**Abstract** We propose a simplified version of the partially observed quasi-information matrix (Poquim) method for inference about non-Gaussian linear mixed models and show its computational advantage over the original method. We illustrate the difference, and compare performance of the simplified version with Poquim as well as the normality-based method in simulation studies. An example of real-data analysis is considered.

**Keywords** asymptotic covariance matrix · Fisher information · Non-Gaussian linear mixed model · quasi-likelihood · REML · Spoquim

## 1 Introduction

Linear mixed models (LMM; e.g., Jiang 2007) are widely used in practice. One advantage that has contributed to the popularity of such models has to do with their robustness, at least in large sample, to violation of the normality assumption, which is often assumed for the distribution of the random effects and errors. In fact, it is known (e.g., Richardson & Welsh 1994, Jiang 1996, 1997) that the normality-based restricted maximum likelihood (REML) estimators of the variance components in a LMM remain consistent and asymptotically normal even if the random effects and errors are not normally distributed; similar results also hold for the maximum likelihood (ML) estimators under suitable restrictions on the dimension of the fixed effects (e.g., Jiang 1996, 1998).

Oregon Health and Science University
Portland, OR, USA

University of California, Davis
Davis, CA, USA
E-mail: jimjiang@ucdavis.edu

In spite of the large-sample properties, inference about non-Gaussian LMM becomes complicated when it comes to estimating the asymptotic covariance matrix (ACM) of the REML or ML estimator because, without the normality assumption, the ACM does not have the well-known form as inverse of the Fisher information matrix; in fact, the ACM involves additional higher-order (third and fourth) moments of the random effects and errors, whose estimates are not available in the outputs of standard software packages, such as SAS and R. To illustrate specifically, let us consider Gaussian REML estimation under a non-Gaussian LMM, which can be regarded as quasi-likelihood estimation (e.g., Heyde 1997). It is known [e.g., Jiang 2007, eq. (1.27)] that the ACM of the (Gaussian) REML estimator has the familiar "sandwich" expression:

$$\Sigma_{\mathrm{R}} = I_2^{-1} I_1 I_2^{-1}, \tag{1}$$

where $I_1 = \mathrm{Var}(\partial l_{\mathrm{R}}/\partial\theta)$, $I_2 = -\mathrm{E}(\partial^2 l_{\mathrm{R}}/\partial\theta\partial\theta')$, $l_{\mathrm{R}}$ is the restricted log-likelihood function, and $\theta$ is the vector of variance components. Note that, here, the expressions of $\partial l_{\mathrm{R}}/\partial\theta$ and $\partial^2 l_{\mathrm{R}}/\partial\theta\partial\theta'$ are derived under the normality assumption, but the expectation and covariance matrix, E and Var, are taken with respect to the true underlying distribution, which may not be Gaussian. Under the normality assumption, one has $I_2 = I_1$, hence (1) becomes $I_1^{-1}$, which is the inverse of the Fisher-information matrix for REML estimation. However, if the normality assumption does not hold, the information identity, $I_2 = I_1$, may not hold. In fact, in such a case, $I_2$ remains the same as that under the normality assumption, which depends only on $\theta$, but $I_1$ involves the fourth moments of the random effects and errors, in additional to the variance components, $\theta$. Similar difference also exists for the ML estimation under a non-Gaussian LMM, in which case the corresponding $I_1$

also involves the third moments of the random effects and errors, in additional to $\theta$ and the fourth moments. Ignoring such a difference in inference may lead to incorrect standard errors, inaccurate confidence intervals and/or misleading assessment of statistical significance.

In order to address such an issue, Jiang (2007, sec. 1.4) proposed two methods of estimating the ACM under a non-Gaussian LMM. The first is called empirical method of moments (EMM). Basically, one obtains estimators of the fourth moments of the random effects and errors, assuming that the third moments of the random effects and errors are zero. One then replaces the fourth moments involved in $I_1$ in (1) by their EMM estimators, and also $\theta$ by $\hat{\theta}$, the REML estimator of $\theta$. This leads to an estimator of $\Sigma_{\mathrm{R}}$. An obvious restriction of EMM is that the third moments of the random effects and errors are zero. This assumption holds if, in particular, the random effects and errors are symmetrically distributed; but, like normality, symmetry also may not hold in practice. The second proposed method, which is the main subject of study of this paper, is called partially observed quasi-information matrix (Poquim). For the most part, one avoids taking expectations of certain terms that potentially leads to the higher moments; instead, those terms are handled in a way similar to the observed information in ML estimation (e.g., Efron & Hinkley 1978). The final expression of the Poquim is a sum of two terms, the first similar to the observed information matrix and the second to the estimated information matrix, hence explaining the name, Poquim. The term quasi-information matrix refers to (1), or a similar expression for ML estimation, when the distribution of the data is non-Gaussian.

The Poquim method does not require the third moments of the random effects and errors being zero; in this regard, the method is more broadly applied than EMM. In the original paper of Poquim, Jiang (2005), the author established some nice theoretical properties of the method. On the other hand, despite the theoretical advance, the method has never been implemented, mainly because of its complex form (see below), which remains a hurdle in the computation. The main purpose of the current paper is to offer a (much) simplified version of Poquim, called simplified Poquim, or Spoquim, and compare its performance with Poquim and EMM via empirical studies. In Section 2, we provide an overview of Poquim and discuss its computational challenge; the Spoquim method is naturally introduced in this context. In Section 3 we make a comparison between Spoquim and Poquim via a simple example, and carry out a Monte-Carlo simulation study that compares the performance of Spoquim, Poquim and EMM. A real-data example is discussed in Section 4. R code for implementing Spoquim is provided in the Supplementary Material.

## 2 Poquim: Computational challenge and remedy

Consider a non-Gaussian linear mixed model that has a general ANOVA structure:

$$y = X\beta + Z_1\alpha_1 + \cdots + Z_s\alpha_s + \epsilon, \qquad (2)$$

where $X, Z_1, \ldots, Z_s$ are known matrices, $\beta$ is an unknown vector of fixed effects, and $\alpha_1, \ldots, \alpha_s, \epsilon$ are independent vectors of random effects and errors such that the components of $\alpha_r$ are i.i.d. with mean 0 and variance $\sigma_r^2, 1 \le r \le s$ and the components of $\epsilon$ are i.i.d. with mean 0 and variance $\tau^2$. Let $m_r = \dim(\alpha_r), 1 \le r \le s$, and $n = \dim(y) = \dim(\epsilon)$. Let $\gamma_r = \sigma_r^2/\tau^2, 1 \le r \le s$, and $\theta = (\tau^2, \gamma_1, \ldots, \gamma_s)' = (\theta_r)_{0\le r\le s}$.

### 2.1 Spoquim for REML

Let us first consider REML estimation. As noted, the focus is on $I_1$. According to Jiang (2005), the Poquim of $I_1$ is given by $\hat{I}_1 = (\hat{I}_{1,qr})_{0\le q,r\le s}$, where $\hat{I}_{1,qr} = \hat{I}_{1,1,qr} + \hat{I}_{1,2,qr}$ and $\hat{I}_{1,a,qr}, a = 1, 2$ are defined below. First, we have

$$\hat{I}_{1,1,qr} = \sum_{f(i_1,i_2,i_3,i_4)\neq 0} \hat{c}_{qr}(i_1, i_2, i_3, i_4)\hat{u}_{i_1}\hat{u}_{i_2}\hat{u}_{i_3}\hat{u}_{i_4} \qquad (3)$$

and the notations in (3) are defined below. Let $u = y - X\beta = (u_i)_{1\le i\le n}$, $u_i$ be the $i$th component of $u$, and $\hat{u}_i$ be $u_i$ with $\beta$ replaced by $\hat{\beta}$, the REML estimator of $\beta$. Next, let

$$f(i_1, i_2, i_3, i_4) = \sum_{r=0}^{s} \kappa_r z_{i_1 r} \cdot z_{i_2 r} \cdot z_{i_3 r} \cdot z_{i_4 r}, \qquad (4)$$

where $\kappa_s = \mathrm{E}(\alpha_{r1}^4) - 3\sigma_r^4, 0 \le r \le s$, with $\alpha_0 = \epsilon$ and $\sigma_0^2 = \tau^2$. Finally, $z'_{ir}$ is the $i$th row of $Z_r, 0 \le r \le s$ with $Z_0 = I_n$, the $n$-dimensional identity matrix, and

$$z_{i_1 r} \cdot z_{i_2 r} \cdot z_{i_3 r} \cdot z_{i_4 r} = \sum_{k=1}^{m_r} z_{i_1 rk} z_{i_2 rk} z_{i_3 rk} z_{i_4 rk}$$

with $z_{irk}$ being the $(i, k)$ element of $Z_r, 1 \le k \le m_r, 0 \le r \le s$, with $m_0 = n$. The $f(i_1, i_2, i_3, i_4)$ defined by (4) is viewed as a function of $\kappa = (\kappa_r)_{0\le r\le s}$; therefore, $f(i_1, i_2, i_3, i_4) \neq 0$ means that the function is not a zero function (which equals to 0 for any $\kappa$). Furthermore, let $f_1, \ldots, f_L$ be the different nonzero functional values of $f(i_1, i_2, i_3, i_4)$, that is, $f_l, 1 \le l \le L$ are all of the different nonzero functions of $\kappa$ resulted from (4) as

$i_1, i_2, i_3, i_4$ vary between 1 and $n$. Let $\hat{c}_{qr}(i_1, i_2, i_3, i_4) = \hat{c}_{qr,l}$ if $f(i_1, i_2, i_3, i_4) = f_l$, where

$$\hat{c}_{qr,l} = \frac{1}{d_r} \sum_{f(i_1,i_2,i_3,i_4)=f_l} \hat{B}_{q,i_1 i_2} \hat{B}_{r,i_3 i_4}, \tag{5}$$

where $d_r = |\{(i_1, i_2, i_3, i_4) : f(i_1, i_2, i_3, i_4) = f_l\}|$ and $|S|$ denotes the cardinality of set $S$, $\hat{B}_{r,i_1 i_2}$ is the $(i_1, i_2)$ element of $\hat{B}_r, 0 \le r \le s$. Here, $B_0 = P/2\tau^2$, $B_r = (\tau^2/2)PZ_rZ_r'P$, $1 \le r \le s$ with

$$P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1},$$

$V = \mathrm{Var}(y) = \tau^2 \sum_{r=0}^{s} \gamma_r Z_r Z_r'$ ($\gamma_0 = 1$), and $\hat{B}_r$ is $B_r$ with $\theta$ replaced by $\hat{\theta}$, the REML estimator of $\theta$. Note that

$$B_r = \frac{1}{2} P \frac{\partial V}{\partial \theta_r} P, \quad 0 \le r \le s. \tag{6}$$

This completes the definition of $\hat{I}_{1,1,qr}$. Next, define $\Gamma(i_1, i_2) = \sum_{r=0}^{s} \gamma_r z_{i_1 r} \cdot z_{i_2 r}$ with

$$z_{i_1 r} \cdot z_{i_2 r} = \sum_{k=1}^{m_r} z_{i_1 rk} z_{i_2 rk},$$

$0 \le r \le s$. Then, we have

$$\begin{aligned}\hat{I}_{1,2,qr} &= 2\mathrm{tr}(\hat{B}_q \hat{V} \hat{B}_r \hat{V}) \\ &\quad -3\hat{\tau}^4 \sum_{f(i_1,i_2,i_3,i_4)\neq 0} \hat{c}_{qr}(i_1, i_2, i_3, i_4) \\ &\quad \times \hat{\Gamma}(i_1, i_3)\hat{\Gamma}(i_2, i_4),\end{aligned} \tag{7}$$

where $\hat{V}$ and $\hat{\Gamma}(i_1, i_2)$ are $V$ and $\Gamma(i_1, i_2)$ with $\theta$ replaced by $\hat{\theta}$, respectively.

The two terms $\hat{I}_{1,1,qr}$ and $\hat{I}_{1,2,qr}$ correspond to the "observed" and "estimated" parts of Poquim, respectively. The main difficulty with Poquim has to do with identifying all different functional values of $f(i_1, i_2, i_3, i_4)$, which, although may seem simple from a mathematical standpoint, is not easy to implement in terms of computer programming. Below we propose an alternative approach that follows the same basic idea of Poquim, but does it in a way that is computationally easier to implement.

According to Jiang (2007, sec. 1.8.5), we have

$$\partial l_{\mathrm{R}} / \partial \theta_r = u' B_r u - b_r$$

with $b_r = \mathrm{E}(u' B_r u)$, $0 \le r \le s$. Thus, we have

$$\begin{aligned}I_{1,qr} &= \mathrm{cov}\left(\frac{\partial l_{\mathrm{R}}}{\partial \theta_q}, \frac{\partial l_{\mathrm{R}}}{\partial \theta_r}\right) = \mathrm{cov}(u' B_q u, u' B_r u) \\ &= \sum_{i_1,i_2,i_3,i_4} B_{q,i_1 i_2} B_{r,i_3 i_4} \mathrm{cov}(u_{i_1} u_{i_2}, u_{i_3} u_{i_4}).\end{aligned} \tag{8}$$

Note that

$$\begin{aligned}&\mathrm{cov}(u_{i_1} u_{i_2}, u_{i_3} u_{i_4}) \\ &= \mathrm{E}(u_{i_1} u_{i_2} u_{i_3} u_{i_4}) - \mathrm{E}(u_{i_1} u_{i_2})\mathrm{E}(u_{i_3} u_{i_4}).\end{aligned}$$

The basic idea of Poquim is not to compute $\mathrm{E}(u_{i_1} u_{i_2} u_{i_3} u_{i_4})$ analytically, if some fourth moments of the random effects and errors are going to appear in the result (the third moments will not appear). Now let us see how the fourth moments are going to show up. According to Lemma 1.3 of Jiang (2007; same as Lemma 1, part 1 of Jiang 2005), we have

$$\begin{aligned}&\mathrm{cov}(u_{i_1} u_{i_2}, u_{i_3} u_{i_4}) \\ &= \tau^4\{\Gamma(i_1, i_3)\Gamma(i_2, i_4) + \Gamma(i_1, i_4)\Gamma(i_2, i_3)\} \\ &\quad + \sum_{r=0}^{s} \kappa_r z_{i_1 r} \cdot z_{i_2 r} \cdot z_{i_3 r} \cdot z_{i_4 r}.\end{aligned} \tag{9}$$

Because $\Gamma$ depends only on $\theta$, the fourth moments will show up if and only if at least one of the 4-factor inner products, $z_{i_1 r} \cdot z_{i_2 r} \cdot z_{i_3 r} \cdot z_{i_4 r}$, is nonzero, which is equivalent to

$$\sum_{r=0}^{s} |z_{i_1 r} \cdot z_{i_2 r} \cdot z_{i_3 r} \cdot z_{i_4 r}| > 0. \tag{10}$$

Let $\mathcal{P}$ denote the subset of indexes $(i_1, i_2, i_3, i_4)$ such that (10) holds. By (8), (9), we have

$$\begin{aligned}&I_{1,qr} \\ &= \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4}\{\mathrm{E}(u_{i_1} u_{i_2} u_{i_3} u_{i_4}) \\ &\quad -\tau^4 \Gamma(i_1, i_2)\Gamma(i_3, i_4)\} \\ &\quad +\tau^4 \sum_{(i_1,i_2,i_3,i_4)\notin\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4}\{\Gamma(i_1, i_3)\Gamma(i_2, i_4) \\ &\quad +\Gamma(i_1, i_4)\Gamma(i_2, i_3)\} \\ &= \mathrm{E}\left\{\sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4} u_{i_1} u_{i_2} u_{i_3} u_{i_4}\right\} \\ &\quad +\tau^4 \left[\sum_{(i_1,i_2,i_3,i_4)\notin\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4}\{\Gamma(i_1, i_3)\Gamma(i_2, i_4) \right. \\ &\quad +\Gamma(i_1, i_4)\Gamma(i_2, i_3)\} \\ &\quad \left. - \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4} \Gamma(i_1, i_2)\Gamma(i_3, i_4)\right],\end{aligned} \tag{11}$$

noting that $\mathrm{E}(u_{i_1} u_{i_2}) = \tau^2 \Gamma(i_1, i_2)$. Furthermore, note that

$$\begin{aligned}&\tau^4 \sum_{(i_1,i_2,i_3,i_4)\notin\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4}\{\Gamma(i_1, i_3)\Gamma(i_2, i_4) \\ &\quad +\Gamma(i_1, i_4)\Gamma(i_2, i_3)\} \\ &= \tau^4 \sum_{i_1,i_2,i_3,i_4} B_{q,i_1 i_2} B_{r,i_3 i_4}\{\Gamma(i_1, i_3)\Gamma(i_2, i_4) \\ &\quad +\Gamma(i_1, i_4)\Gamma(i_2, i_3)\} \\ &\quad -\tau^4 \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4}\{\Gamma(i_1, i_3)\Gamma(i_2, i_4) \\ &\quad +\Gamma(i_1, i_4)\Gamma(i_2, i_3)\}.\end{aligned} \tag{12}$$

For the first term on the right side of (2), note that, in view of (9), this is $\text{cov}(u'B_q u, u'B_r u)$ when the random effects and errors are Gaussian, which implies that the kurtoses $\kappa_r, 0 \leq r \leq s$ in (9) are all zero. It follows that the first term on the right side of (12) is equal to

$$2\text{tr}(B_q V B_r V) = \frac{1}{2}\text{tr}\left(P\frac{\partial V}{\partial \theta_q}P\frac{\partial V}{\partial \theta_r}\right), \quad (13)$$

using an identity for the covariance of normal quadratic forms (e.g., Jiang 2007, p. 38), (6), and the identity $PVP = P$. Note that $\partial V/\partial \tau^2 = V/\tau^2$ and $\partial V/\partial \gamma_r = \tau^2 Z_r Z_r', 1 \leq r \leq s$. As for the second term on the right side of (12), note that $(i_1, i_2, i_3, i_4) \in \mathcal{P}$ if and only if $(i_1, i_2, i_4, i_3) \in \mathcal{P}$, and $B_{r,i_3 i_4} = B_{r,i_4 i_3}$. It follows that the second term on the right side of (12) is equal to $-2\tau^4 \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4} \Gamma(i_1, i_3)\Gamma(i_2, i_4)$. Combining these simplifications, we can write (11) further as $I_{1,qr} = I_{1,1,qr} + I_{1,2,qr}$, where

$$I_{1,1,qr} = \text{E}\left\{\sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} B_{q,i_1 i_2} B_{r,i_3 i_4} u_{i_1} u_{i_2} u_{i_3} u_{i_4}\right\}, \quad (14)$$

$$I_{1,2,qr} = \frac{1}{2}\text{tr}\left(P\frac{\partial V}{\partial \theta_q}P\frac{\partial V}{\partial \theta_r}\right) - \tau^4 \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} B_{q,i_1 i_1} B_{r,i_3 i_4}\{\Gamma(i_1, i_2)\Gamma(i_3, i_4) + 2\Gamma(i_1, i_3)\Gamma(i_2, i_4)\}. \quad (15)$$

The $I_{1,1,qr}$ in (14) corresponds to the "observed" and the $I_{1,2,qr}$ in (15) the "estimated" parts, respectively. Following the Poquim idea, for the observed part, we remove the expectation sign in (14), and replace the unknown parameters involved in $B$ and $u$ by their REML estimators. This leads to

$$\tilde{I}_{1,1,qr} = \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} \hat{B}_{q,i_1 i_2} \hat{B}_{r,i_3 i_4} \hat{u}_{i_1} \hat{u}_{i_2} \hat{u}_{i_3} \hat{u}_{i_4}. \quad (16)$$

For the estimated part, which only depends on $\theta$, we simply replace $\theta$ by $\hat{\theta}$. This leads to

$$\tilde{I}_{1,2,qr} = 2\text{tr}(\hat{B}_q \hat{V} \hat{B}_r \hat{V}) - \hat{\tau}^4 \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} \hat{B}_{q,i_1 i_2} \hat{B}_{r,i_3 i_4}\{\hat{\Gamma}(i_1, i_2)\hat{\Gamma}(i_3, i_4) + 2\hat{\Gamma}(i_1, i_3)\hat{\Gamma}(i_2, i_4)\}, \quad (17)$$

using (13). Note that the first term on the right side of (17) is the same as that on the right side of (7). The combination of (16) and (17) leads to the simplified Poquim (Spoquim):

$$\tilde{I}_{1,qr} = \tilde{I}_{1,1,qr} + \tilde{I}_{1,2,qr}, \quad 0 \leq q, r \leq s, \quad (18)$$

and $\tilde{I}_1 = (\tilde{I}_{1,qr})_{0\leq q,r\leq s}$. The main simplicity of Spoquim over Poquim is in that the functions $f(i_1, i_2, i_3, i_4)$ and the quantities $c_{qr}(i_1, i_2, i_3, i_4)$ involved in (3) and (7), whose computations are not straightforward, are avoided altogether.

On the other hand, the $I_2$ in (1) has the same expression as that under normality, that is, $I_2 = (I_{2,qr})_{0\leq q,r\leq s}$, where $I_{2,qr}$ is equal to (13). Because $I_2$ depends only on $\theta$, an estimator is obtained by replacing the $\theta$ by $\hat{\theta}$. Denote the resulting estimator of $I_2$ by $\hat{I}_2$. An estimator of the ACM, $\Sigma_\text{R}$ in (1), is then given by $\hat{\Sigma}_\text{R} = \hat{I}_2^{-1}\tilde{I}_1\hat{I}_2^{-1}$.

## 2.2 Spoquim for ML

Let $l$ denote the Gaussian log-likelihood function. According to Jiang (2005, sec. 4), we have $\partial l/\partial \beta = X'V^{-1}u$, $\partial l/\partial \theta_r = u'C_r u - c_r$, where $C_0 = V^{-1}/2\tau^2$, $C_r = (\tau^2/2)V^{-1}Z_r Z_r' V^{-1}, 1 \leq r \leq s$, and $c_r = \text{E}(u'C_r u), 0 \leq r \leq s$.The quasi-information matrix has a similar expression as (1), that is,

$$\Sigma = \mathcal{I}_2^{-1}\mathcal{I}_1\mathcal{I}_2^{-1}, \quad (19)$$

where $\mathcal{I}_1 = \text{Var}(\partial l/\partial \psi)$ and $\mathcal{I}_2 = \text{E}(\partial^2 l/\partial \psi \partial \psi')$ with $\psi = (\beta', \theta')'$. Furthermore, $\mathcal{I}_2$ has the same expression as that under the normality assumption, that is [e.g., Jiang 2007, eq. (1.13)–(1.15)], we have

$$\mathcal{I}_2 = \text{diag}\{\partial^2 l/\partial \beta \partial \beta', \text{E}(\partial^2 l/\partial \theta \partial \theta')\} \text{ with}$$

$$\frac{\partial^2 l}{\partial \beta \partial \beta'} = -X'V^{-1}X,$$

$$\text{E}\left(\frac{\partial^2 l}{\partial \theta_q \partial \theta_r}\right) = -\frac{1}{2}\text{tr}\left(V^{-1}\frac{\partial V}{\partial \theta_q}V^{-1}\frac{\partial V}{\partial \theta_r}\right), \quad (20)$$

$0 \leq q, r \leq s$. Thus, the focus is on estimation of $\mathcal{I}_1$. We have (e.g., Jiang 2005)

$$\text{Var}\left(\frac{\partial l}{\partial \beta}\right) = X'V^{-1}X, \quad (21)$$

which is easily estimated by replacing $V$ by $\hat{V}$, which is obtained by replacing $\theta$ in the expression of $V$ by its ML estimator. Furthermore, the Spoquim of $\text{Var}(\partial l/\partial \theta)$ is the same way as that of $I_1$ for REML estimation, with the only difference being replacing $B$ by $C$, that is, (16), (17) with $B$ replaced by $C$. Finally, we have

$$\text{cov}\left(\frac{\partial l}{\partial \beta_j}, \frac{\partial l}{\partial \theta_r}\right) = \text{cov}(X_j'V^{-1}u, u'C_r u) = X_j'V^{-1}\left[\sum_{i_2,i_3} C_{r,i_2 i_3}\text{cov}(u_{i_1}, u_{i_2}u_{i_3})\right]_{1\leq i_1\leq n}, \quad (22)$$

where $X_j$ is the $j$th column of $X$, $1 \leq j \leq p = \dim(\beta)$, $0 \leq r \leq s$. According to Jiang (2005; Lemma 2), we have

$$\text{cov}(u_{i_1}, u_{i_2}u_{i_3}) = \sum_{r=0}^{s} \text{E}(\alpha_{r1}^3) z_{i_1 r} \cdot z_{i_2 r} \cdot z_{i_3 r}, \quad (23)$$

where $z_{i_1r} \cdot z_{i_2r} \cdot z_{i_3r} = \sum_{k=1}^{m_r} z_{i_1rk} z_{i_2rk} z_{i_3rk}$. Thus, at least one third moment of random effects or errors will appear in (23) provided that

$$\sum_{r=0}^{s} |z_{i_1r} \cdot z_{i_2r} \cdot z_{i_3r}| > 0. \tag{24}$$

Let $\mathcal{Q}$ denote the subset of indexes $(i_1, i_2, i_3)$ such that (24) holds, and $H_i$ the $i$th column of $V^{-1}, 1 \le i \le n$. Then, by (22), (23), we have

$$\begin{aligned} &\text{cov}\left(\frac{\partial l}{\partial \beta_j}, \frac{\partial l}{\partial \theta_r}\right) \\ &= \sum_{i_1,i_2,i_3} X_j' H_{i_1} C_{r,i_2i_3} \text{cov}(u_{i_1}, u_{i_2}u_{i_3}) \\ &= \sum_{(i_1,i_2,i_3)\in\mathcal{Q}} X_j' H_{i_1} C_{r,i_2i_3} \text{E}(u_{i_1}u_{i_2}u_{i_3}) \\ &= \text{E}\left\{ \sum_{(i_1,i_2,i_3)\in\mathcal{Q}} X_j' H_{i_1} C_{r,i_2i_3} u_{i_1}u_{i_2}u_{i_3} \right\}. \end{aligned} \tag{25}$$

It follows that the covariance consists only of the observed part, which is estimated by

$$\begin{aligned} &\widetilde{\text{cov}}\left(\frac{\partial l}{\partial \beta_j}, \frac{\partial l}{\partial \theta_r}\right) \\ &= \sum_{(i_1,i_2,i_3)\in\mathcal{Q}} X_j' \hat{H}_{i_1} \hat{C}_{r,i_2i_3} \hat{u}_{i_1}\hat{u}_{i_2}\hat{u}_{i_3}, \quad 1 \le j \le p, \end{aligned} \tag{26}$$

$0 \le r \le s$, where $\hat{H}_i$ is the $i$th column of $\hat{V}^{-1}$ and $\hat{C}_r$ is $C_r$ with $\theta$ replaced by its ML estimator, $\hat{\theta}$.

## 3 Example, simulation, and computing notes

We first use a simple example to illustrate Spoquim and its difference with Poquim. We then carry out a simulation study under the same example to compare the performances of Spoquim and Poquim. We conclude the section with some notes on computing and coding.

### 3.1 Example

Jiang (2005) used a simple example of balanced one-way random effects model to illustrate Poquim. The model can be expressed as

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \tag{27}$$

$i = 1, \dots, m, j = 1, \dots, b$, where $\mu$ is an unknown mean, $\alpha_i$ is a random effect and $\epsilon_{ij}$ is an error. It is assumed that the random effects and errors are independent with mean 0, and $\text{var}(\alpha_i) = \sigma^2$, $\text{var}(\epsilon_{ij}) = \tau^2$. The model can be expressed as (2) with $X = 1_m \otimes 1_b$, $s = 1$, and $Z_1 = Z = I_m \otimes 1_b$, where $\otimes$ denotes Kronecker product.

We focus on REML estimation. It can be shown that

$$\begin{aligned} B_0 &= \frac{1}{2\tau^4}\left(I_m \otimes I_b - \frac{\sigma^2}{\tau^2 + b\sigma^2} I_m \otimes J_b\right) \\ &\quad - \frac{\bar{J}_m \otimes \bar{J}_b}{2\tau^2(\tau^2 + b\sigma^2)}, \end{aligned} \tag{28}$$

$$B_1 = \frac{\tau^2}{2(\tau^2 + b\sigma^2)^2}(I_m - \bar{J}_m) \otimes J_b. \tag{29}$$

where $\bar{J}_k = k^{-1} J_k$ with $J_k$ being the $k \times k$ matrix of 1s. It is more convenient to use a double index, $(i, j)$, with $1 \le i \le m, 1 \le j \le b$ than a single index, $i$, with $1 \le i \le n = mb$ when relevant. For example, $y = (y_1, \dots, y_n)'$ is the same as

$$y = (y_{11}, \dots, y_{1b}, y_{21}, \dots, y_{2b}, \dots, y_{m1}, \dots, y_{mb})'.$$

Then, the $(i, j)$th row of $Z_0 = I_n = I_m \otimes I_b$ is the row vector $z'_{(i,j)0} = [1_{\{(k,l)=(i,j)\}}]'_{1\le k\le m, 1\le l\le b}$, that is, the $1 \times n$ vector whose $(i, j)$ component is 1 and other components are 0. Thus, we have

$$\begin{aligned} &z_{(i_1,j_1)0} \cdot z_{(i_2,j_2)0} \cdot z_{(i_3,j_3)0} \cdot z_{(i_4,j_4)0} \\ &= \sum_{k=1}^{m}\sum_{l=1}^{b} 1_{\{(k,l)=(i_1,j_1)\}} 1_{\{(k,l)=(i_2,j_2)\}} \\ &\quad \times 1_{\{(k,l)=(i_3,j_3)\}} 1_{\{(k,l)=(i_4,j_4)\}}. \end{aligned}$$

which is positive if and only if $(i_1, j_1) = (i_2, j_2) = (i_3, j_3) = (i_4, j_4)$, that is, $i_1 = i_2 = i_3 = i_4$ and $j_1 = j_2 = j_3 = j_4$. Also, the $(i, j)$ row of $Z_1 = I_m \otimes 1_b$ is the row vector $z_{(i,j)1k} = [1_{(k=i)}]_{1\le k\le m}$, that is, the $1 \times m$ vector whose $i$th component is 1 and other components are 0. It follows that

$$\begin{aligned} &z_{(i_1,j_1)1} \cdot z_{(i_2,j_2)1} \cdot z_{(i_3,j_3)1} \cdot z_{(i_4,j_4)1} \\ &= \sum_{k=1}^{m} 1_{(k=i_1)} 1_{(k=i_2)} 1_{(k=i_3)} 1_{(k=i_4)}, \end{aligned}$$

which is positive if and only if $i_1 = i_2 = i_3 = i_4$. It follows that the index set $\mathcal{P}$ defined below (10) can be expressed as

$$\begin{aligned} \mathcal{P} = \{&[(i_1, j_1), (i_2, j_2), (i_3, j_3), (i_4, j_4)] : \\ &i_1 = i_2 = i_3 = i_4\}. \end{aligned} \tag{30}$$

By (28), the $[(i_1, j_1), (i_2, j_2)]$ element of $B_0$ is

$$\begin{aligned} &B_{0,(i_1,j_1)(i_2,j_2)} \\ &= \frac{1}{2\tau^4}\left\{1_{(i_1=i_2,j_1=j_2)} - \frac{\sigma^2}{\tau^2 + b\sigma^2} 1_{(i_1=i_2)}\right\} \\ &\quad - \frac{(mb)^{-1}}{2\tau^2(\tau^2 + b\sigma^2)}. \end{aligned}$$

Similarly, by (29), the $[(i_1, j_1), (i_2, j_2)]$ element of $B_1$ is

$$B_{1,(i_1,j_1)(i_2,j_2)} = \frac{\tau^2}{2(\tau^2 + b\sigma^2)^2}\left\{1_{(i_1=i_2)} - \frac{1}{m}\right\}.$$

Using these expressions, and (30), it can be derived that

$$\tilde{I}_{1,1,00} = \frac{1}{4\hat{\tau}^8}\sum_{i=1}^m \left\{\sum_{j=1}^b \hat{u}_{ij}^2 - \frac{\hat{\sigma}^2+\hat{\tau}^2/n}{\hat{\tau}^2+b\hat{\sigma}^2}\left(\sum_{j=1}^b \hat{u}_{ij}\right)^2\right\}^2,$$

$$\tilde{I}_{1,1,01} = \frac{1-1/m}{4\hat{\tau}^2(\hat{\tau}^2+b\hat{\sigma}^2)^2}\sum_{i=1}^m \left\{\sum_{j=1}^b \hat{u}_{ij}^2 - \frac{\hat{\sigma}^2+\hat{\tau}^2/n}{\hat{\tau}^2+b\hat{\sigma}^2}\left(\sum_{j=1}^b \hat{u}_{ij}\right)^2\right\}\left(\sum_{j=1}^b \hat{u}_{ij}\right)^2,$$

$$\tilde{I}_{1,1,11} = \frac{\hat{\tau}^4(1-1/m)^2}{4(\hat{\tau}^2+b\hat{\sigma}^2)^4}\sum_{i=1}^m \left(\sum_{j=1}^b \hat{u}_{ij}\right)^4,$$

where $\hat{u}_{ij} = y_{ij} - \hat{\mu}$ and $\hat{\mu}$ is the REML estimator of $\mu$. Similarly, it can be derived that

$$\tilde{I}_{1,2,00} = \frac{m}{2\hat{\tau}^4}\left\{1 - \frac{1}{m} - \left(1-\frac{1}{m}\right)^2 - \frac{b^2}{2}\left(1 - \frac{\hat{\sigma}^2/m+\hat{\tau}^2/n}{\hat{\tau}^2+b\hat{\sigma}^2}\right)^2\right\},$$

$$\tilde{I}_{1,2,01} = \frac{(m-1)b}{2(\hat{\tau}^2+b\hat{\sigma}^2)}\left\{\frac{1}{m} - \frac{b}{2}\left(1 - \frac{\hat{\sigma}^2/m+\hat{\tau}^2/n}{\hat{\tau}^2+b\hat{\sigma}^2}\right)\right\},$$

$$\tilde{I}_{1,2,11} = \frac{(m-1)b^2\hat{\tau}^4}{4(\hat{\tau}^2+b\hat{\sigma}^2)^2}\left(\frac{3}{m}-1\right).$$

Jiang (2005; also see Jiang 2007, p. 47) derived detailed expressions of Poquim for this example. It can be seen that $f(i_1j_1,\dots,i_4j_4) = 0$, if not $i_1 = \cdots = i_4$; $\kappa_1$, if $i_1 = \cdots = i_4$ but not $j_1 = \cdots = j_4$; and $\kappa_0+\kappa_1$, if $i_1 = \cdots = i_4$ and $j_1 = \cdots j_4$. Thus, $L = 2$ [note that $L$ is the number of different functional values of $f(i_1j_1,\dots,i_4j_4)$]. Define the following functions of $\theta = (\tau^2,\gamma)'$ with $\gamma = \sigma^2/\tau^2$: $t_0 = 1-\gamma/(1+b\gamma) - 1/\{(1+b\gamma)n\}$ (note that $n = mb$), $t_1 = (m-1)b/\{m(1+b\gamma)\}$, and $t_2 = \{b(1+b\gamma)^2-(1+\gamma)^2\}/(b^3-1)$. Then, the Poquim is given by $\hat{\mathcal{I}}_{1,qr} = \hat{\mathcal{I}}_{1,1,qr}+\hat{\mathcal{I}}_{1,2,qr}$, $q,r = 0,1$, where

$$\hat{\mathcal{I}}_{1,1,00} = \frac{\hat{t}_1^2-\hat{t}_0^2 b}{4\hat{\tau}^8 b(b^3-1)}\left\{\sum_{i=1}^m\left(\sum_{j=1}^b \hat{u}_{ij}\right)^4 - \sum_{i=1}^m\sum_{j=1}^b \hat{u}_{ij}^4\right\} + \frac{\hat{t}_0^2}{4\hat{\tau}^8}\sum_{i=1}^m\sum_{j=1}^b \hat{u}_{ij}^4,$$

**Table 1** Values of Spoquim (SPOQ) and Poquim (POQ) from Example Data

| | | | |
|---|---|---|---|
| SPOQ | $\tilde{I}_{1,1,00} = 919.6722$ | $\tilde{I}_{1,2,00} = -430.0872$ | $\tilde{I}_{1,00} = 489.585$ |
| POQ | $\hat{I}_{1,1,00} = 742.989$ | $\hat{I}_{1,2,00} = -422.0935$ | $\hat{I}_{1,00} = 320.8955$ |
| SPOQ | $\tilde{I}_{1,1,01} = 141.09$ | $\tilde{I}_{1,2,01} = -47.47662$ | $\tilde{I}_{1,01} = 93.61338$ |
| POQ | $\hat{I}_{1,1,01} = 128.8025$ | $\hat{I}_{1,2,01} = -18.45532$ | $\hat{I}_{1,01} = 110.3472$ |
| SPOQ | $\tilde{I}_{1,1,11} = 67.53872$ | $\tilde{I}_{1,2,11} = -5.10062$ | $\tilde{I}_{1,11} = 62.4381$ |
| POQ | $\hat{I}_{1,1,11} = 67.53872$ | $\hat{I}_{1,2,11} = -5.10062$ | $\hat{I}_{1,11} = 62.4381$ |

$$\hat{\mathcal{I}}_{1,1,01} = \frac{(m-1)(\hat{t}_1 b-\hat{t}_0)}{4\hat{\tau}^6(1+b\hat{\gamma})^2 m(b^3-1)}\left\{\sum_{i=1}^m\left(\sum_{j=1}^b \hat{u}_{ij}\right)^4 - \sum_{i=1}^m\sum_{j=1}^b \hat{u}_{ij}^4\right\} + \frac{(m-1)\hat{t}_0}{4\hat{\tau}^6(1+b\hat{\gamma})^2 m}\sum_{i=1}^m\sum_{j=1}^b \hat{u}_{ij}^4,$$

$$\hat{\mathcal{I}}_{1,1,11} = \frac{(m-1)^2}{4\hat{\tau}^4(1+b\hat{\gamma})^4 m^2}\sum_{i=1}^m\left(\sum_j \hat{u}_{ij}\right)^4;$$

$$\hat{\mathcal{I}}_{1,2,00} = \frac{1}{2\hat{\tau}^4}\left(mb-1 - \frac{3}{2}m[b\hat{t}_0^2\{(1+\hat{\gamma})^2-\hat{t}_2\}+\hat{t}_1^2\hat{t}_2]\right),$$

$$\hat{\mathcal{I}}_{1,2,01} = \frac{(m-1)b}{2\hat{\tau}^2(1+b\hat{\gamma})}\left\{1 - \left(\frac{3}{2}\right)\frac{(b\hat{t}_1-\hat{t}_0)\hat{t}_2+(1+\hat{\gamma})^2\hat{t}_0}{1+b\hat{\gamma}}\right\},$$

$$\hat{\mathcal{I}}_{1,2,11} = -\frac{(m-1)(m-3)b^2}{4m(1+b\hat{\gamma})^2},$$

where $\hat{\gamma} = \hat{\sigma}^2/\hat{\tau}^2$, and the $\hat{t}$'s are the $t$'s with $\theta$ replaced by $\hat{\theta}$, the REML estimator.

It is seen that, compared to Spoquim, the expressions of Poquim are more complicated involving the functions $t_a, a = 0,1,2$, with the exception of $\hat{I}_{1,a,11}, a = 1,2$, which are equal to their Spoquim counterparts. In order to see the difference, we compute $\hat{I}_{1,a,qr}$ and $\tilde{I}_{1,a,qr}$, $(q,r) \neq (1,1), a = 1,2$ numerically. A dataset is generated under model (27) with $\mu = \sigma^2 = \tau^2 = 1$, $m = 50$ and $b = 5$. The distributions of the random effects and errors are centralized exponential with rate equal to 1, that is, the distribution of $\xi - 1$, where $\xi \sim$ Exponential(1). This distribution is denoted by CE(1). Note that the mean and variance of CE(1) are 0 and 1, respectively. The REML estimates of $\mu, \sigma^2, \tau^2$ are 0.930, 1.105 and 0.848, respectively. Based on the data and REML estimates, the Spoquim and Poquim are computed and summarized on Table 1.

### 3.2 Simulation study

From the previous subsection we see that the values of Spoquim and Poquim may be different. A simulation study is carried out under the same model of the previous subsection to compare the performance of Spoquim and Poquim in estimating the ACM, given by (1). As noted, $I_2$ is the same as that under normality, which, in this case, is given by

$$I_{2,00} = \frac{n-1}{2\tau^4},$$
$$I_{2,01} = \frac{(m-1)b}{2(\tau^2 + b\sigma^2)},$$
$$I_{2,11} = \frac{\tau^4 (m-1)b^2}{2(\tau^2 + b\sigma^2)^2}, \qquad (31)$$

where $n = mb$. The difference between Spoquim and Poquim in estimating the ACM is therefore in that $\tilde{\Sigma}_{\rm R} = \hat{I}_2^{-1}\tilde{I}_1\hat{I}_2^{-1}$ for Spoquim and $\hat{\Sigma}_{\rm R} = \hat{I}_2^{-1}\hat{I}_1\hat{I}_2^{-2}$ for Poquim, where the expressions of $\tilde{I}_1, \hat{I}_1$ are given in the previous subsection, and $\hat{I}_2$ is given by (31) with $\theta$ replaced by $\hat{\theta}$, the REML estimator.

The simulation setting covers normality and varies situations of non-normality, including heavy-tail, skewed, and bimodal distributions. Specifically, there are four cases: (I) both $\alpha$ and $\epsilon$ are normal; (II) both $\alpha$ and $\epsilon$ are $\sqrt{2/3}t_6$; (III) both $\alpha$ and $\epsilon$ are CE(1) (see the previous subsection); and (IV) both $\alpha$ and $\epsilon$ are equal-weight mixture of $N(1/\sqrt{2}, 1/2)$ and $N(-1/\sqrt{2}, 1/2)$. We also consider a case of discrete random effects and errors, namely, (V) $\alpha$ has equal probability mass of 0.5 at $-1$ and 1 and $\epsilon$ has probability masses of 1/12, 1/6, 1/2, 1/6, 1/12 at $-2$, $-1$, 0, 1, 2, respectively. It is easy to see that all of the distributions have been standardized to have mean 0 and variance 1, and finite fourth moment. Two performance measures of ACM estimation are considered. The first is percentage relative bias (%RB), defined as

$$\%\text{RB} = 100 \times \left\{ \frac{\text{E}(\hat{\sigma}_{qr}) - \sigma_{qr}}{\sigma_{qr}} \right\}, \quad 0 \le q \le r \le 1,$$

where $\hat{\sigma}_{qr}$ is the $(q,r)$ element of the estimator of $\Sigma_{\rm R}$, either by Spoquim or by Poquim, $\text{E}(\hat{\sigma}_{qr})$ is the simulated mean of $\hat{\sigma}_{qr}$, and $\sigma_{qr}$ is the $(q,r)$ element of the simulated true covariance matrix, $\text{Var}(\hat{\theta})$. The second is coefficient of variation (CV), defined as

$$\text{CV} = \frac{\sqrt{\text{var}(\hat{\sigma}_{qr})}}{|\text{E}(\hat{\sigma}_{qr})|}, \quad 0 \le q \le r \le 1.$$

The results, based on 1,000 simulation runs, are presented in Table 2 It is seen that the overall performance of Spoquim and Poquim are quite close. In terms of absolute value of %RB, Spoquim performs better in exactly half of the cases (6 out of 12); in terms of CV, Spoquim performs better in 8 out of the 12 cases. Considering that Spoquim is much easier to program to a software package (see Appendix) than Poquim, we therefore conclude that there is a significant computational advantage of Spoquim over Poquim without losing estimation efficiency.

**Table 2** Comparing Performance of Spoquim (SPOQ) and Poquim (POQ) in Terms of %RB and CV

| | | | %RB | | CV | |
|---|---|---|---|---|---|---|
| Case | Element | True $\sigma_{qr}$ | SPOQ | POQ | SPOQ | POQ |
| (I) | (0,0) | 0.010 | 1.05 | -3.40 | 0.31 | 0.45 |
| | (0,1) | -0.011 | 8.77 | 0.33 | 0.40 | 0.43 |
| | (1,1) | 0.069 | 4.48 | 1.88 | 0.50 | 0.49 |
| (II) | (0,0) | 0.023 | -4.70 | -8.58 | 1.38 | 1.34 |
| | (0,1) | -0.025 | -11.72 | -15.28 | 0.87 | 1.00 |
| | (1,1) | 0.139 | -6.43 | -7.11 | 1.51 | 1.57 |
| (III) | (0,0) | 0.034 | 0.66 | 4.85 | 0.80 | 0.98 |
| | (0,1) | -0.040 | -20.48 | -10.76 | 0.73 | 1.03 |
| | (1,1) | 0.254 | -12.60 | -7.84 | 1.78 | 1.80 |
| (IV) | (0,0) | 0.007 | 6.67 | 3.05 | 0.27 | 0.43 |
| | (0,1) | -0.010 | -1.45 | -10.41 | 0.41 | 0.37 |
| | (1,1) | 0.066 | -7.52 | -10.64 | 0.51 | 0.47 |
| (V) | (0,0) | 0.010 | -0.47 | -2.87 | 0.25 | 0.22 |
| | (0,1) | -0.013 | -2.77 | -7.93 | 0.26 | 0.19 |
| | (1,1) | 0.035 | 0.09 | -3.40 | 0.36 | 0.35 |

**Table 3** Comparing Spoquim (SPOQ), Poquim (POQ) and Normality-based Method (NBM) in Terms of %RB and Size (Nominal Level = 0.10)

| Sample Size | Case | %RB | | | Size | | |
|---|---|---|---|---|---|---|---|
| | | SPOQ | POQ | NBM | SPOQ | POQ | NBM |
| $m = 50$ | (I) | -5.00 | -7.26 | 1.85 | 0.138 | 0.146 | 0.114 |
| $b = 5$ | (II) | -6.00 | -6.67 | -45.87 | 0.151 | 0.158 | 0.201 |
| | (V) | -1.37 | -5.26 | 116.41 | 0.109 | 0.113 | 0.028 |
| $m = 100$ | (I) | -8.90 | -9.72 | -4.25 | 0.138 | 0.134 | 0.112 |
| $b = 5$ | (II) | -2.99 | -3.05 | -45.71 | 0.125 | 0.128 | 0.217 |
| | (V) | -0.38 | -2.21 | 120.93 | 0.101 | 0.112 | 0.015 |

It may be wondered what impact one may get by ignoring the possible non-normality, and use the normality-based method to obtain the variance of the REML estimator. This issue was previously addressed in Jiang (2005), but here we illustrate numerically using the current simulated example. We consider estimation of $\gamma = \sigma^2/\tau^2$ and also testing the hypothesis $\text{H}_0 : \gamma = 1$. We first compare the %RB of the normality method with Poquim and Spoquim. Next, we compare the size (or observed level of significance) for testing $\text{H}_0$ at the 10% level of significance. We make the comparison under normality, case (I), and two non-normal situations, namely, case (II) and case (V). Furthermore, we consider the same sample size configuration, $m = 50, b = 5$, as well as an increased sample size configuration, $m = 100, b = 5$, and observe the changes when the sample size increases. The results, based on 1,000 simulation runs, are presented in Table 3.

It is seen that, under normality [case (I)], NBM performs better both in terms of %RB and in terms of the size; however, the trend is reversed, significantly, under the non-normal situations [case (II) and case (V)]. Also note that, in terms of %BR, although NBM performs better under normality, the %RB for all three methods stay within single-digit, which is generally considered good performance. On the other hand, in the non-normal situations the %RB for NBM is much higher in the range of mid double-digit or low triple-digit, which are considered poor performance; meanwhile, the %RB of SPOQ and POQ remain single-digit. Finally, it is observed that, the performance of SPOQ and POQ, in terms of the size, improves as the sample size increases. However, this does not seem to be the case for NBM; in fact, the size for NBM seems to get a little worse as the sample size increases in the non-normal cases.

An inaccurate estimate of the variance of the estimator, whose square root is the standard error (s.e.), or inaccurate level of significance, may be misleading in practice. Thus, if there is a serious doubt about the normality of the random effects or errors, which is often the case in practice, one should use Spoquim or Poquim in assessing uncertainty, and Spoquim is computationally more attractive than Poquim as we have shown.

### 3.3 Some notes on computing

The expression of Spoquim, (16) and (17), involve summation of a 4-way array over a subset of indexes $1 \le i_1, i_2, i_3, i_4 \le n$, where $n$ is the sample size. A brute-force coding that involves a 4-way loop is computationally inefficient and extremely slow even for moderately large $n$. The following are some computing ideas that avoid the 4-way loop and, as a result, speed up the computation tremendously.

First consider the index set $\mathcal{P}$ defined by (10). Define $a * b = (a_k b_k)_{1\le k\le K}$ for $a = (a_k)_{1\le k\le K}$ and $b = (b_k)_{1\le k\le K}$. Recall $z'_{ir}$ is the $i$th row of $Z_r$. Then, for fixed $(i_1, i_2)$,

$$z_{i_1 r}\cdot z_{i_2 r}\cdot z_{i_3 r}\cdot z_{i_4 r} = \sum_{k=1}^{m_r} z_{i_1 rk} z_{i_2 rk} z_{i_3 rk} z_{i_4 rk}$$

is the $(i_3, i_4)$ element of $Z_r \mathrm{diag}(z_{i_1 r} * z_{i_2 r}) Z'_r$. It follows that, for fixed $i_1, i_2$, $\sum_{r=0}^{s} |z_{i_1 r}\cdot z_{i_2 r}\cdot z_{i_3 r}\cdot z_{i_4 r}|$ is the $(i_3, i_4)$ element of the matrix $S = \sum_{r=0}^{s} |Z_r \mathrm{diag}(z_{i_1 r} * z_{i_2 r}) Z'_r|$. Here $|A| = (|a_{ij}|)$ for matrix $A = (a_{ij})$ [i.e., the abs() function in R]. The R code ifelse($S > 0, 1, 0$) then returns a matrix $Q = Q_{i_1 i_2}$, whose elements are 1 if (10) holds, and 0 otherwise.

Now consider expression (16). The expression can be written as

$$\begin{aligned} S_1 &= \sum_{i_1,i_2} \left\{ \sum_{(i_3,i_4):(i_1,i_2,i_3,i_4)\in\mathcal{P}} \hat{B}_{r,i_3 i_4} \hat{u}_{i_3} \hat{u}_{i_4} \right\} \\ &\quad \times \hat{B}_{q,i_1 i_2} \hat{u}_{i_1} \hat{u}_{i_2} \\ &= \sum_{i_1,i_2} \hat{u}'[\hat{B}_{q,i_1 i_2} \hat{u}'[\hat{B}_{r,i_3 i_4} Q_{i_1 i_2, i_3 i_4}]_{1\le i_3,i_4\le n} \hat{u} \\ &\quad ]_{1\le i_1,i_2\le n} \hat{u} \\ &= \hat{u}' \left[ \hat{B}_{q,i_1 i_2} \hat{u}' (\hat{B}_r * Q_{i_1 i_2}) \hat{u} \right] \hat{u}, \qquad (32) \end{aligned}$$

where $Q_{i_1 i_2, i_3 i_4}$ is the $(i_3, i_4)$ element of $Q_{i_1 i_2}$, and the $*$ operation between two matrices is defined similarly as that between two vectors, that is, $A * B = (a_{ij} b_{ij})$ for matrices $A = (a_{ij})$ and $B = (b_{ij})$. Next, we consider (17). There are two main 4-way summations here, namely,

$$S_{2,1} = \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} \hat{B}_{q,i_1 i_2} \hat{B}_{r,i_3 i_4} \hat{\Gamma}(i_1, i_2) \hat{\Gamma}(i_3, i_4),$$

$$S_{2,2} = \sum_{(i_1,i_2,i_3,i_4)\in\mathcal{P}} \hat{B}_{q,i_1 i_2} \hat{B}_{r,i_3 i_4} \hat{\Gamma}(i_1, i_3) \hat{\Gamma}(i_2, i_4).$$

The first can be written as

$$\begin{aligned} S_{2,1} &= \sum_{i_1,i_2} \left\{ \sum_{i_3,i_4} \hat{B}_{r,i_3 i_4} \hat{\Gamma}(i_3, i_4) Q_{i_1 i_2, i_3 i_4} \right\} \\ &\quad \times \hat{B}_{q,i_1 i_2} \hat{\Gamma}(i_1, i_2) \\ &= \sum_{i_1,i_2} \mathrm{sum}(\hat{B}_r * \hat{\Gamma} * Q_{i_1 i_2}) \hat{B}_{q,i_1 i_2} \hat{\Gamma}(i_1, i_2) \\ &= \mathrm{sum}([\mathrm{sum}(\hat{B}_r * \hat{\Gamma} * Q_{i_1 i_2})]_{1\le i_1,i_2\le n} \\ &\quad * \hat{B}_q * \hat{\Gamma}), \qquad (33) \end{aligned}$$

where sum() denotes the sum function in R,

$$\hat{\Gamma} = [\hat{\Gamma}(i_1, i_2)]_{1\le i_1,i_2\le n},$$

and the $*$ operation for three matrices is defined similarly as for two. The second can be written as

$$\begin{aligned} S_{2,2} &= \sum_{i_1,i_2} \left\{ \sum_{i_3,i_4} \hat{B}_{r,i_3 i_4} Q_{i_1 i_2, i_3 i_4} \hat{\Gamma}(i_1, i_3) \hat{\Gamma}(i_2, i_4) \right\} \\ &\quad \times \hat{B}_{q,i_1 i_2} \\ &= \sum_{i_1,i_2} \hat{\Gamma}(i_1)' (\hat{B}_r * Q_{i_1 i_2}) \hat{\Gamma}(i_2) \hat{B}_{q,i_1 i_2} \\ &= \mathrm{sum}([\hat{\Gamma}(i_1)' (\hat{B}_r * Q_{i_1 i_2}) \hat{\Gamma}(i_2)]_{1\le i_1,i_2\le n} \\ &\quad * \hat{B}_q), \qquad (34) \end{aligned}$$

where $\hat{\Gamma}(i)'$ denotes the $i$th row of $\hat{\Gamma}$.

With (32)–(34), efficient R codes can be written (see Section A.1 of the Supplementary Material) to compute the Spoquim for REML, $S = S_1 + 2\mathrm{tr}(\hat{B}_q \hat{V} \hat{B}_r \hat{V}) - \hat{\tau}^4 \sum_{a=1}^{2} a S_{2,a}$ [see (18), (16), (17)]. Spoquim for ML codes can be developed similarly (see Appendix A.2 of the Supplementary Material).

For large $n$, the idea of parallel computing can be used. Note that each of the expressions (32)–(34) involves a matrix whose indexes are $1 \le i_1, i_2 \le n$. The matrix can be partitioned into sub-matrices, and computation of the sub-matrices can be assigned to different computers. This strategy can be used to compute Spoquim when $n$ is large.

## 4 Real-data example

As a real-data example, we consider the Television School and Family Smoking Prevention and Cessation Project

(TVSFP; e.g., Hedeker *et al.* 1994). The original study was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation. The subjects were seventh-grade students from schools in Los Angeles (LA) and San Diego in the State of California in the United States. The students were pretested in January 1986 in an initial study. The same students completed an immediate post-intervention questionnaire in April 1986, a one-year follow-up questionnaire (in April 1987), and a two-year follow-up (in April 1988). Schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum (CC); (b) a media (television) intervention (TV); (c) a combination of CC and TV conditions; and (d) a no-treatment control. A tobacco and health knowledge scale (THKS) score was one of the primary study outcome variables. The THKS consisted of seven questionnaire items used to assess student tobacco and health knowledge. A student's THKS score was defined as the sum of the items that the student answered correctly. The data is available at www.hsph.harvard.edu/fitzmaur/ala/tvsfp.txt.

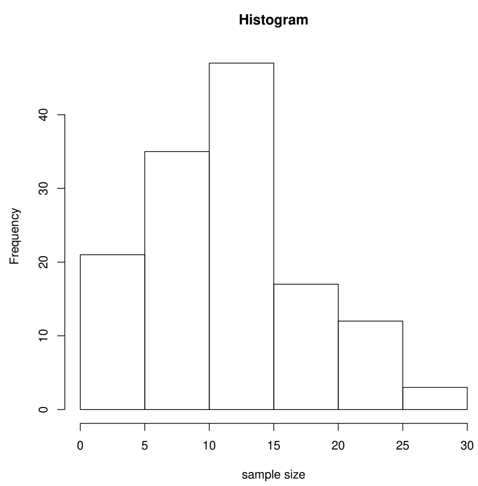


**Fig. 1** Histogram of Sample Sizes from Classes in LA Schools

Here, we consider part of the data involving $m = 28$ LA schools. The schools are considered as the clusters associated with the random effects. There were $n = 135$ classes within those schools that participated in the study, or a total of 1,600 student. The sample sizes from those classes ranged from 1 to 28. A histogram of the sample sizes is presented in Figure 1. For each of those students, the THKS scores were available for the pretest and immediate post-intervention studies. The response variable that we consider, is the average mean difference in THKS score, that is, the immediate post-intervention score minus the pretest score for each sampled student, averaged over the class.

Hedeker *et al.* (1994) considered fitting the TVSFP data using normality-based maximum likelihood. Note that the THKS score is integer-valued ranging from 0 to 7. Also, due to the relatively small sample size from each class, the central limit theorem does not apply to those averages. As a result, the response variable is clearly not normal. Thus, we consider fitting a Non-Gaussian LMM using REML and Spoquim to address the non-normality issue. The LMM can be expressed as

$$y_{ij} = \beta_0 + \beta_1 \mathrm{CC}_{ij} + \beta_2 \mathrm{TV}_{ij} + \beta_3 \mathrm{CCTV}_{ij} + \alpha_i + \epsilon_{ij}, \quad (35)$$

$i = 1, \dots, 28$, $j = 1, \dots, b_i$, where $\mathrm{CC}_{ij}, \mathrm{TV}_{ij}$ are indicators on whether the class participated in the CC program, or TV program, and $\mathrm{CCTV}_{ij}$ is the product of $\mathrm{CC}_{ij}$ and $\mathrm{TV}_{ij}$ representing the interaction. Here $\beta_k, k = 0, 1, 2, 3$ are unknown fixed effects, $\alpha_i$ is a school-level random effect, and $\epsilon_{ij}$ an additional error. The random effects and errors are assumed to be independent such that $\alpha_i \sim N(0, \sigma^2), \epsilon_{ij} \sim N(0, \tau^2)$, where $\sigma^2, \tau^2$ are unknown variances. Here, $\tau^2$ and $\gamma = \sigma^2/\tau^2$ are considered as the variance components.

The REML estimates for the fixed effects are $\hat{\beta}_0 = 0.269$, $\hat{\beta}_1 = 0.632$, $\hat{\beta}_2 = 0.050$ and $\hat{\beta}_3 = -0.105$. The standard errors for the fixed-effect estimates are obtained from the standard outputs after fitting the REML with R, namely, $\mathrm{s.e.}(\hat{\beta}_0) = 0.038$, $\mathrm{s.e.}(\hat{\beta}_1) = 0.054$, $\mathrm{s.e.}(\hat{\beta}_2) = 0.054$ and $\mathrm{s.e.}(\hat{\beta}_3) = 0.078$. Thus, one may interpret the result as that the intercept and CC indicator are significant, while the TV indicator and CCTV interaction are not significant (say, at 5% level of significance).

Furthermore, the REML estimates of the variance components are $\hat{\tau}^2 = 0.050$ and $\hat{\gamma} = 12.610$. Due to the non-normality of the data, the normality-based Fisher information matrix may not be appropriate to be used to obtain the standard errors (see Section 1). Therefore, we compute the Spoquim using the R code that we have developed. This leads to the following results (see the end of Section 2.1):

$$\hat{I}_1 = \begin{pmatrix} 1.494 \times 10^7 & 42.756 \\ 42.756 & 0.037 \end{pmatrix},$$

$$\hat{I}_2 = \begin{pmatrix} 26040.534 & 2.193 \\ 2.193 & 0.008 \end{pmatrix},$$

$$\hat{\Sigma}_{\mathrm{R}} = \hat{I}_2^{-1} \hat{I}_1 \hat{I}_2^{-1} = \begin{pmatrix} 0.023 & -6.107 \\ -6.107 & 2180.161 \end{pmatrix}.$$

It follows that the standard errors for $\hat{\tau}^2$ and $\hat{\gamma}$ are 0.152 and 46.692, respectively.

A main purpose of this real-data example is to illustrate a situation where it may not be appropriate to use the normality-based method to compute the standard errors (s.e.) associated with the variance component estimates; the Spoquim method is more appropriate to use. As noted earlier, the data are clearly not normal; therefore, the normality-based s.e. is expected to be less accurate than the Spoquim s.e. (Jiang 2005; also see our additional result in Section 3.2), even though the normality-based REML estimators may still be appropriate to use as point estimators (Jiang 1996, 1997).

## 5 Declarations

There are no financial or non-financial interests regarding any of the coauthors.
**Acknowledgement:** The research of Thuan Nguyen and Jiming Jiang are partially supported by the National Science Foundation of the United States grants DMS-1914760 and DMS-1914465, respectively. Jiming Jiang's research is also supported by the National Science Foundation grant DMS-1713120. The authors are grateful to the reviewers' comments that have helped improve the work and presentation.

**Description of contents in the Supplementary Material:** The Supplementary Material contains R codes for REML and ML estimation under a general mixed ANOVA model. The codes have been verified in special cases, where closed-form expressions of the Spoquim can be derived, to make sure that the outcomes are the same. An example is given to help the users get familiar with the codes.

## References

1. Efron, B. and Hinkley, D. V. (1978), Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information, *Biometrika* 65, 457–487.
2. Hedeker, D., Gibbons, R. D., and Flay, B. R. (1994), Random-effects regression models for clustered data with an example from smoking prevention research, *J. Consulting Clinical Psych.* 62, 757-765.
3. Heyde, C. C. (1997), *Quasi-likelihood and Its Application*, Springer, New York.
4. Jiang, J. (1996), REML estimation: Asymptotic behavior and related topics, *Ann. Statist.* 24, 255–286.
5. Jiang, J. (1997), Wald consistency and the method of sieves in REML estimation, *Ann. Statist.* 25, 1781–1803.
6. Jiang, J. (1998), Asymptotic properties of the empirical BLUP and BLUE in mixed linear models, *Statistica Sinica* 8, 861–885.
7. Jiang, J. (2005), Partially observed information and inference about non-Gaussian mixed linear models, *Ann. Statist.* 33, 2695–2731.
8. Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, Springer, New York.
9. Richardson, A. M. and Welsh, A. H. (1994), Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models, *Austral. J. Statist.* 36, 31–43.