

Open Journal of Statistics, 2015, 5, 568-584

Published Online October 2015 in SciRes. <http://www.scirp.org/journal/ojs>

<http://dx.doi.org/10.4236/ojs.2015.56059>



Statistical Inference in Generalized Linear Mixed Models by Joint Modelling Mean and Covariance of Non-Normal Random Effects

Yin Chen¹, Yu Fei², Jianxin Pan^{3*}

¹School of Insurance, University of International Business and Economics, Beijing, China

²School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, China

³School of Mathematics, University of Manchester, Manchester, UK

Email: chenyin@uibe.edu.cn, feiyukm@aliyun.com, jianxin.pan@manchester.ac.uk

Received 8 September 2015; accepted 20 October 2015; published 23 October 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Generalized linear mixed models (GLMMs) are typically constructed by incorporating random effects into the linear predictor. The random effects are usually assumed to be normally distributed with mean zero and variance-covariance identity matrix. In this paper, we propose to release random effects to non-normal distributions and discuss how to model the mean and covariance structures in GLMMs simultaneously. Parameter estimation is solved by using Quasi-Monte Carlo (QMC) method through iterative Newton-Raphson (NR) algorithm very well in terms of accuracy and stabilization, which is demonstrated by real binary salamander mating data analysis and simulation studies.

Keywords

Generalized Linear Mixed Models, Multivariate t Distribution, Multivariate Mixture Normal Distribution, Quasi-Monte Carlo, Newton-Raphson, Joint Modelling of Mean and Covariance

1. Introduction

Generalized linear mixed models (GLMMs) are very helpful and widely used for analyzing discrete data and data from exponential family distributions. Statistical inference of GLMMs is challenging due to the incorporation of random effects, especially for example, the marginal likelihood has the form of analytical intractable high dimensional integration. The existing and popular statistical methods include: 1) analytical Laplace approximation: the uncorrelated penalized quasi-likelihood (PQL) by Breslow and Clayton (1993) [1] and correlated PQL by Lin and Breslow (1996) [2]; hierarchical generalized linear models (HGLM) procedure by Lee and

*Corresponding author.

How to cite this paper: Chen, Y., Fei, Y. and Pan, J.X. (2015) Statistical Inference in Generalized Linear Mixed Models by Joint Modelling Mean and Covariance of Non-Normal Random Effects. *Open Journal of Statistics*, 5, 568-584.

<http://dx.doi.org/10.4236/ojs.2015.56059>

Nelder (2001) [3]; 2) numerical technique: Bayesian approach with sampling by Karim and Zeger (1992) [4]; MCEM algorithm by Booth and Hobert (1999) [5]; Gauss-Hermite quadrature (GHQ) by Pan and Thompson (2003) [6]; Quasi-Monte Carlo (QMC) and Randomized QMC method by Pan and Thompson (2007) [7] and Aleid (2007) [8].

One common idea in above literatures is random effects in GLMMs, which represent latent effects between individuals, follow normal distribution with mean zero and identity covariance matrix. However, these assumptions are not always valid in practice because individual has its own special natures and may not be normal distributed. To address this issue, multivariate t distribution and multivariate mixture normal distribution will be assumed for random effects in GLMMs. Quasi-Monte Carlo (QMC) method through iterative Newton-Raphson (NR) algorithm solves the high-dimensional integration problem in the marginal likelihood under the non-normal assumptions.

Review of GLMMs and the marginal Quasi-likelihood will be proposed in Section 2. QMC approximation, the simplest Good Point set (square root sequence) and analytical form of maximum likelihood estimates (MLEs) in GLMMs will be discussed in Section 3. The idea of modified cholesky decomposition and joint modelling of mean and covariance structure will be explained in Section 4. In Section 5, the salamander data will be analyzed as an example under the assumption of multivariate t distribution of random effects, then simulation study with same design protocol as salamander data followed in Section 6. Discussions on the related issues and further studies are given in Section 7.

2. Generalized Linear Mixed Models

The generalized linear mixed models (GLMMs) are typically constructed by incorporating random effects into the linear predictor of a conditionally independent exponential family model (McCulloch, 2003, [9]).

$$\begin{aligned}
 Y_i | \mathbf{b} &\sim f_{Y_i|\mathbf{b}}(y_i | \mathbf{b}); \\
 f_{Y_i|\mathbf{b}}(y_i | \mathbf{b}) &= \exp\left[\frac{y_i \mathcal{G}_i - d(\mathcal{G}_i)}{a_i(\varphi)} + c(y_i, \varphi)\right]; \\
 E(Y_i | \mathbf{b} = \mathbf{b}) &= \mu_i^{\mathbf{b}}; \\
 g(\mu_i^{\mathbf{b}}) &= \eta_i^{\mathbf{b}} = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{b}; \\
 \mathbf{b} &\sim f_{\mathbf{b}}(\mathbf{b});
 \end{aligned}$$

where Y_i denotes the observed response, $i=1,2,\dots,n$, \mathbf{x}'_i is the i^{th} row of design matrix; $\boldsymbol{\beta}$ is $p \times 1$ fixed effect parameter; \mathbf{z}'_i is the i^{th} row of the second design matrix; \mathbf{b} is $q \times 1$ random effect.

$\mu_i^{\mathbf{b}} = E(Y_i | \mathbf{b} = \mathbf{b}) = d'(\mathcal{G}_i)$ and $\text{var}(Y_i | \mathbf{b} = \mathbf{b}) = \frac{\varphi}{a_i} d''(\mathcal{G}_i) = \frac{\varphi}{a_i} v(\mu_i^{\mathbf{b}})$, where a_i is a prior weight, φ is a

known overdispersion scalar, \mathcal{G}_i is the natural parameter, and $v(\cdot)$ is variance function (McCulloch and Nelder, 1989, [10]).

In this definition we see the usual ingredients of a generalized linear model. First, the distribution of Y_i from an exponential family (in this case the distribution is assumed to hold conditional on the random effect \mathbf{b}). Second, a link function, $g(\cdot)$ is applied to the conditional mean of Y_i given \mathbf{b} to obtain the conditional linear predictor $\eta_i^{\mathbf{b}}$. When $\eta_i^{\mathbf{b}}$ is identical to the \mathcal{G}_i , GLMMs is said to have canonical link function. Finally, the linear predictor is assumed to consist of two components, the fixed effects portion, described by $\mathbf{x}'_i \boldsymbol{\beta}$ and random effects portion, $\mathbf{z}'_i \mathbf{b}$, for which a distribution is assigned to \mathbf{b} . e.g. if \mathbf{b} is said to follow a q -dimensional t distribution, denote by $\mathbf{b} \sim T_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, and its density function is given by

$$f_{\mathbf{b}}(\mathbf{b}) = \frac{\Gamma\left(\frac{\nu+q}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{\frac{q}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left(1 + \frac{(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+q}{2}} \quad (1)$$

with the mean of \mathbf{b} is $\mu (\nu > 1)$ and the covariance matrix of \mathbf{b} is $\frac{\nu \Sigma}{\nu - 2} (\nu > 2)$ respectively, where μ is a $q \times 1$ vector, Σ is a $q \times q$ symmetric positive definite matrix, ν is the degree of freedom. If \mathbf{b} is said to follow a q -dimensional mixture normal distribution F

$$F = \pi N_q[-(1-\pi)\lambda \mathbf{1}_q, \Sigma] + (1-\pi) N_q(\pi \lambda \mathbf{1}_q, \Sigma) \tag{2}$$

with mean $0_{q \times 1}$, covariance matrix $\Sigma + \pi(1-\pi)\mathbf{1}_q \mathbf{1}'_q$ where $\pi \in [0, 1]$, $\mathbf{1}_q = (1, \dots, 1)'$ is $q \times 1$ vector, Σ is a $q \times q$ symmetric positive definite matrix.

In GLMMs, statistical inferences of β and θ are estimated in most literatures, β is $p \times 1$ fixed effect parameter, θ is a vector parameter in covariance matrix of random effect, $\Sigma_b(\theta)$. Quasi-likelihood function of (β, θ) in GLMMs is also called the *marginal quasi-likelihood function* which expressed by

$$L(\beta, \theta) = \exp\{\ell(\beta, \theta)\} = \int \exp\left\{\sum_{i=1}^n \ell_i(\beta, \theta | \mathbf{b})\right\} dF(b; \theta), \tag{3}$$

where

$$\ell_i(\beta, \theta | \mathbf{b}) \propto \int_{y_i}^{\mu_i} \frac{a_i(y_i - u)}{\phi_\nu(u)} du, \tag{4}$$

$$\ell(\beta, \theta) = \log L(\beta, \theta) = \log \int \exp\left\{\sum_{i=1}^n \ell_i(\beta, \theta | \mathbf{b})\right\} dF(b; \theta),$$

determinates the *conditional log quasi-likelihood* of β and θ given \mathbf{b} (Breslow and Clayton, 1993, [1]). $F(b; \theta)$ is cumulative distribution function (CDF) of random effect b .

It is extremely challenging to obtain the maximum likelihood estimates (MLEs), $(\hat{\beta}, \hat{\theta})$, which maximize (3) or (4). Although the integration involves an analytical expression, the computational problems are magnified when the specified model contains a large number of random effects and random effects have a crossed designed according to data description.

3. Quasi-Monte Carlo Integration and Estimation

3.1. Quasi-Monte Carlo Integration

Quasi-Monte Carlo (QMC) sequences are a deterministic alternative to Monte Carlo (MC) sequences (Niederreiter, 1992, [11]). For the q -dimensional unit cube domain $C^q = [0, 1]^q$, the Quasi-Monte Carlo approximation is

$$\int_{C^q} f(x) dx \approx \frac{1}{K} \sum_{k=1}^K f(x_k),$$

where sample points x_1, x_2, \dots, x_K should be selected to minimize the error of the integral quadrature.

If the integrand f has finite variation, then the error of the quasi-Monte Carlo quadrature can be bounded using the Koksma-Hlawka inequality

$$\left| \int_{C^q} f(x) dx - \frac{1}{K} \sum_{k=1}^K f(x_k) \right| \leq V(f) D^*(x_1, x_2, \dots, x_K), \tag{5}$$

where $V(f)$ is variation of $f(x)$ in the sense of Hardy and Krause (Fang and Wang, 1994, [12]) and $D^*(x_1, x_2, \dots, x_K)$ is the star discrepancy of sample points, which is a measure of uniformity of distribution of a finite point set $\{x_k : 1 \leq k \leq K\}$, defined by

$$D^*(x_1, x_2, \dots, x_K) = \sup_A \left| \frac{m(A)}{K} - V(A) \right|,$$

where A is an arbitrary q -dimensional subcube parallel to the coordinate axes and originating at the centre, $V(A)$ is its volume, and $m(A)$ is the number of sample points inside this subcube.

For carefully selected sample points, the discrepancy and consequently the error of low-discrepancy sequ-

ences (Niederreiter, 1992, [11]) can be in the order of $O\left((\log K)^{q-1} K^{-1}\right)$, which is much better than $O\left(K^{-1/2}\right)$ the probabilistic error bound of Monte Carlo methods (Fang and Wang, 1994, [12]). Moreover, quasi-Monte Carlo methods guarantee this accuracy in a deterministic way, unlike Monte Carlo methods where the error bound is also probabilistic.

In the QMC approach, there exist precise construction algorithms for generating the required points. These algorithms can be divided into two families, the *Lattice rule* and *Digital-net*. Only the good point set, which is belong to lattice rule (Fang and Wang, 1994, [12]) will be used and discussed in following sections.

3.2. The Good Point (GP) Set

A good point $\nu = (\nu_1, \nu_2, \dots, \nu_q) \in C^q$, where $C^q = [0,1]^q$ is a hypercube. The set P_k consists of the first K points of the set $\left\{\left\{\nu_1 k\right\},\left\{\nu_2 k\right\}, \dots,\left\{\nu_q k\right\}\right\}, k=1,2, \dots, K$, where $\{y\}$ represents the fractional part of the value y . In practice, the following forms of the good point ν are recommended as square root sequence

$$\nu = \left(\left\{\sqrt{p_1}\right\},\left\{\sqrt{p_2}\right\}, \dots,\left\{\sqrt{p_q}\right\}\right),$$

where $p_i, 1 \leq i \leq q$, is a series of prime numbers, for example, the first q primes $p_i \neq p_j$ if $i \neq j$.

3.3. MLE in GLMMs by QMC Estimation

When QMC method applied to *marginal quasi-likelihood* function (4)

$$\ell(\beta, \theta) = \log \left[\frac{1}{K} \sum_{k=1}^K \exp \left\{ \sum_{i=1}^n \ell_i \left(\beta, \theta \mid \Sigma_b^{\frac{1}{2}}(\theta) F^{-1}\left(c_k\right) \right) \right\} \right], \tag{6}$$

where $F^{-1}(\cdot)$ is the inverse of CDF of \mathbf{b} and $\Sigma_b^{\frac{1}{2}}(\theta)$ is the square root of Σ_b , for example, it can be taken as the Cholesky decomposition of Σ_b or eigenvalue-eigenvector decomposition.

Let $b_k = \Sigma_b^{\frac{1}{2}} F^{-1}\left(c_k\right)$, $\eta_{ik}^{b_k} = x_i^T \beta + z_i^T b_k = x_i^T \beta + z_i^T \Sigma_b^{\frac{1}{2}} F^{-1}\left(c_k\right)$, $\left\{c_k : k=1,2, \dots, K\right\}$ is a good point set $\mu_{ik}^{b_k} = h\left(\eta_{ik}^{b_k}\right)$, $h(\cdot)$ is the inverse function of $g(\cdot)$. The global maximum of $\ell(\beta, \theta)$ with respect to β and θ is the solution of the score equations

$$\frac{\partial \ell(\beta, \theta)}{\partial \beta} = \sum_{k=1}^K w_k \left(\sum_{i=1}^n \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right) \dot{g}\left(\mu_{ik}^{b_k}\right)} \cdot x_i \right) = 0, \tag{7}$$

$$\frac{\partial \ell(\beta, \theta)}{\partial \theta} = \sum_{k=1}^K w_k \left(\sum_{i=1}^n \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right) \dot{g}\left(\mu_{ik}^{b_k}\right)} \cdot z_i^T \Sigma_{\theta_j}^{\frac{1}{2}} F^{-1}\left(c_k\right) \right) = 0, \tag{8}$$

where $\Sigma_{\theta_j}^{\frac{1}{2}} = \frac{\partial \Sigma_b^{\frac{1}{2}}}{\partial \theta_j}$, w_k is a weight for the k th point in the i th subject and it has the following form

$$w_k = \frac{\exp \left\{ \sum_{i=1}^n \ell_i\left(\beta, \theta \mid b_k\right) \right\}}{\sum_{k=1}^K \exp \left\{ \sum_{i=1}^n \ell_i\left(\beta, \theta \mid b_k\right) \right\}}. \tag{9}$$

The weight w_k given in (9) is a function of β and θ , i.e. $w_k \equiv w_k(\beta, \theta)$, which must be taken into account when calculating the second-order derivatives.

The score Equations (7) and (8) in general have no analytical solutions for β and θ . A numeral solution is given by Pan and Thompson (2007) [7], which used Newton-Raphson algorithm.

$$\begin{pmatrix} \beta \\ \theta \end{pmatrix}_{(m+1)} = \begin{pmatrix} \beta \\ \theta \end{pmatrix}_{(m)} + \begin{pmatrix} \frac{\partial^2 \ell}{\partial \beta \partial \beta'} & \frac{\partial^2 \ell}{\partial \beta \partial \theta'} \\ \frac{\partial^2 \ell}{\partial \theta \partial \beta'} & \frac{\partial^2 \ell}{\partial \theta \partial \theta'} \end{pmatrix}_{(m)}^{-1} \begin{pmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial \theta} \end{pmatrix}_{(m)}, \tag{10}$$

where $\frac{\partial \ell}{\partial \beta}$ and $\frac{\partial \ell}{\partial \theta}$ are score function of log likelihood, $\begin{pmatrix} \frac{\partial^2 \ell}{\partial \beta \partial \beta'} & \frac{\partial^2 \ell}{\partial \beta \partial \theta'} \\ \frac{\partial^2 \ell}{\partial \theta \partial \beta'} & \frac{\partial^2 \ell}{\partial \theta \partial \theta'} \end{pmatrix}_{(m)}$ is Hessian matrix of log likelihood.

Calculation of the Hessian matrix is difficult and its computation is intensive. When the difference between $\beta^{(m)}$ and $\beta^{(m-1)}$ and the difference between $\theta^{(m)}$ and $\theta^{(m-1)}$ are both sufficient small, the maximum likelihood estimates are confirmed, that is, the convergence is reached. The calculation process of Hessian matrix is given in Appendix A, which is an extension of Hessian matrix in Pan and Thompson (2007) [7]. At convergence, the MLEs, $(\hat{\beta}, \hat{\theta})'$ and asymptotic variance-covariance matrix of $(\hat{\beta}, \hat{\theta})'$ can be obtained.

4. Modified Cholesky Decomposition and Covariance Modelling for Random Effects

Modified Cholesky decomposition of Σ_b^{-1} , rather than Σ_b , and joint modelling of mean and covariance matrix were published by Pourahmadi (1999, 2000) [13] [14], to obtain a statistically meaningful unconstrained parameterization of covariance. The remarkable advantages are: 1) it guarantees covariance matrix is positive definite; 2) it reduces the number of parameters so that it makes computation efficient; 3) it has very clear statistical interpretation. The main idea of modified cholesky decomposition is that a $q \times q$ symmetric matrix Σ_b is positive definite if and only if there exist a unique unit lower triangular matrix T , with 1 as diagonal entries, and unique diagonal matrix D with positive diagonal entries such that

$$T \Sigma_b T' = D,$$

where

$$T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{21} & 1 & 0 & \cdots & 0 \\ -\phi_{31} & -\phi_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{q1} & -\phi_{q2} & -\phi_{q3} & \cdots & 1 \end{bmatrix} \text{ and } D = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_q^2 \end{bmatrix}.$$

It offers a simple unconstrained and statistically meaningful reparametrization of the covariance matrix. So $\Sigma_b = T^{-1} D (T')^{-1}$ and $\Sigma_b^{-1} = T' D^{-1} T$. T and D are easy to compute and interpret statistically: the below-diagonal entries of T are the negatives of the autoregressive coefficients (ACs), ϕ_{jk} , in the autoregressive model.

$$\hat{b}_j = \sum_{k=1}^{j-1} \phi_{jk} b_k \quad (1 \leq k < j \leq q), \tag{11}$$

where b_j is a 1×1 number, which index the j th dimension of random effect b . That is, the linear least squares predictor of b_j based on its predecessors b_{j-1}, \dots, b_1 . On the other hand, the diagonal entries of D are the innovation variances (IVs), $\sigma_j^2 = \text{Var}(e_j)$, where $e_j = b_j - \hat{b}_j$ with \hat{b}_j given by (11). Obviously we have $\text{Cov}(e_j, e_k) = 0$ if $j \neq k$ ($1 \leq j, k \leq q$). Denote $e = (e_1, \dots, e_q)'$. It is obvious $e = T b$, so that $e \sim N_q(0, D)$.

The linear joint models of the autoregressive coefficients (ACs) and the logarithms of innovation variances (IVs)

$$\begin{cases} \phi_{jk} = c'_{jk}\gamma \\ \log \sigma_j^2 = h'_j\lambda \end{cases} \tag{12}$$

where $c'_{jk} (1 \times q_1^*)$ and $h'_j (1 \times q_2^*)$ are covariates, and γ and λ are low-dimensional parameter vectors. The choice of covariate vectors c_{jk} and h_j are flexible in some senses.

5. Salamander Data Analysis

The famous salamander interbreeding data set was published by McCullagh and Nelder (1989), which came from three repeated same protocol design experiments. The salamanders originally lived in two geographically isolated populations, Roughbutt (RB) and Whiteside (WS). Each experiment involved in two closed groups and each group involved 5 females and 5 males from each population. In a fixed period, each female mated with 6 males, 3 from each population. So 60 correlated binary observations were created in each closed group, and totally, 360 binary observations (1 for successful interbreeding, 0 for failed interbreeding) were in three experiments. The primary two objectives of experiment are to study whether the successful mating rate is significant between populations and if heterogeneity between individuals in the mating probability exists. This binary data set is challenging because it is block crossed, correlated, balanced (as each female mated with a total of six males and vice verse) and incomplete (as each female mated with just three out of a possible five males, and vice verse).

In a closed group, 5 females and 5 males from each population involved, so total 20 salamander were indexed by i or j . Denoting by y_{ij} the outcome for the mating of the i th female with the j th male, the logit mixed model for each experiment

$$\text{logit}(p_{ij}^b) = \log\left(\frac{p_{ij}^b}{1-p_{ij}^b}\right) = x'_{ij}\beta + b_i^f + b_j^m \quad (i, j = 1, \dots, 20), \tag{13}$$

where the conditional probability $p_{ij}^b = \Pr\{y_{ij} = 1 \mid b_i^f, b_j^m\}$ is used to model the correlated binary responses.

$$x_{ij} = (1, ws_i^f, ws_j^m, ws_i^f \times ws_j^m)^T, \tag{14}$$

$ws_i^f = 1$ if the i^{th} female comes from WS, otherwise 0. Similarly, $ws_j^m = 1$ if the j^{th} male comes from WS, otherwise 0.

Here the random effect is not assumed as normal distribution but multivariate t distribution with degree of freedom $\nu = 3, 7, 10, 15$, $\mathbf{b} \sim T_{20}(0, \Sigma, \nu)$.

Note that each experiment includes two closed groups and involves 40 salamanders so that the log-likelihood for each experiment involves a 40-dimensional integrals. The dimensionality for each experiment can be further reduced to the sum of two 20-dimensional integrals due to the block design of two closed groups, see Karim and Zeger (1992) [4] and Shun (1997) [15] for more details about the design. When the three experiments data are pooled as in model (13), the log-likelihood ℓ is a sum of six 20-dimensional integrals so that the MLEs of the fixed effects β and variance components θ become extremely difficult to obtain.

Modelling Σ_b for salamander mating data, $q = 20$, $q_1^* = 4$ and $q_2^* = 4$. The autoregressive coefficients (ACs) model is reordered as

$$\begin{aligned} \Phi &= \begin{pmatrix} \phi_{21} \\ \phi_{31} \\ \phi_{32} \\ \vdots \\ \phi_{20,1} \\ \phi_{20,2} \\ \vdots \\ \phi_{20,19} \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{172} \\ \phi_{173} \\ \vdots \\ \phi_{190} \end{pmatrix} = \begin{pmatrix} c'_1\gamma \\ c'_2\gamma \\ c'_3\gamma \\ \vdots \\ c'_{172}\gamma \\ c'_{173}\gamma \\ \vdots \\ c'_{190}\gamma \end{pmatrix} = \begin{pmatrix} c'_{2,1}\gamma \\ c'_{3,1}\gamma \\ c'_{3,2}\gamma \\ \vdots \\ c'_{20,1}\gamma \\ c'_{20,2}\gamma \\ \vdots \\ c'_{20,19}\gamma \end{pmatrix} = \mathbf{C}\gamma, \\ &= (\phi_{d_1, d_2}) = (\phi_i) = (c'_i\gamma) = (c'_{d_1, d_2}\gamma) \end{aligned}$$

where Φ is a 190×1 vector, C is a 190×4 matrix, c_i' is the i^{th} row of C , $i = 1, \dots, 190$.

$$c_i' = c'_{d_1, d_2} = (1, G_{d_1, d_2}, D_{d_1, d_2}, G_{d_1, d_2} \times D_{d_1, d_2}),$$

where

$$i = 0 + 1 + 2 + \dots + (d_1 - 2) + d_2,$$

$G_{d_1, d_2} = 1$ if d_1^{th} and d_2^{th} salamander come from different genders, otherwise 0. Similarly, $D_{d_1, d_2} = 1$ if d_1^{th} and d_2^{th} salamander come from different districts (populations), otherwise 0.

The logarithms of innovation variances (IVs) model is

$$\Omega = \begin{pmatrix} \log \sigma_1^2 \\ \vdots \\ \log \sigma_{20}^2 \end{pmatrix} = \begin{pmatrix} h_1' \\ \vdots \\ h_{20}' \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = H\lambda$$

where H is a 20×3 matrix, λ is a 3×1 vector, σ_j^2 ($j = 1, \dots, 20$) describes the variance of the j^{th} salamander

$$h_j' = (1, \text{Gender}_j, \text{District}_j) \quad j = 1, \dots, 20,$$

where $\text{Gender}_j = 1$, if the j^{th} salamander is female, otherwise 0. Similarly, $\text{District}_j = 1$ if the j^{th} salamander comes from WS, otherwise 0.

In the logarithms of innovation variances (IVs) model, originally there were 4 covariates including the item $\text{Gender}_j \times \text{District}_j$ with interaction parameter λ_4 . But the results showed that λ_4 approximately equals zero and has smaller order of magnitude than λ_2 and λ_3 . The hypothesis test also showed that there is no evidence to reject $\lambda_4 = 0$.

The salamander data is now analyzed by using QMC approach to calculate the MLEs of the fixed effects and variance components involved in the model. Specifically, we implement the simplest QMC point, square root sequences to estimate the parameters for the modelling of the pooled data. A set of 20-dimensional points on the unit cube $C^{20} = [0, 1]^{20}$ were generated for the six 20-dimensional integrals. Then the points are modified through using transformation, $\Sigma_b^{\frac{1}{2}} F^{-1}(\cdot)$, where $F(\cdot)$ is the cumulative distribution function of the multi-

variate t distribution, $\Sigma_b^{\frac{1}{2}}$ is the modified Cholesky decomposition for $\Sigma_b = \Sigma_b^{\frac{1}{2}} \cdot \left(\Sigma_b^{\frac{1}{2}}\right)'$. After that, we approximate the integrated log-likelihood, then we use the Newton-Raphson algorithm (10) to maximize the approximated log-likelihood function.

The covariance modelling for $\Sigma_b = \text{Cov}(b)$, the covariance matrix of random effect, uses 4 parameters for autoregressive coefficients of b_i and b_j , $1 \leq i, j \leq 20$, (i and j index the dimension of b while b_i means the i^{th} salamander's latent effect) and plus 3 parameters for innovation variances of $b_j - \hat{b}_j$, $\text{var}(b_j - \hat{b}_j)$, $j = 1, \dots, 20$. Using this joint modelling method to analyze the real world data can estimate Σ_b more to the true one because of no factitious assumption on it, instead, only the data tells us what is Σ_b . Another advantage is the parameter is reduced from $\frac{20 \times (20 + 1)}{2}$ to 7 in Σ_b . This reduction of number of parameters makes the computation easier. Note that, the higher the dimension q , the more significant this advantage. Third, the modified Cholesky decomposition confirms that unconstrained parameters lead to strict condition that is the Σ_b is a symmetric positive definite matrix.

In the calculations, the convergence accuracy is set to be $\sqrt{(\beta^{(m)} - \beta^{(m-1)})^2 + (\theta^{(m)} - \theta^{(m-1)})^2} \leq 0.01$. The numerical results from **Tables 1-4** show that for each of the degree of freedom, QMC method make the maximum log-likelihood similar when increasing the number of QMC points from 10,000 to 100,000 which indicates any number of points between these could prove reasonable estimate of the parameters. In addition, the parameter estimates using these sequences become stable quickly. Bearing in mind that the integral space is

Table 1. MLEs of the parameters by covariance modelling for the pooled salamander data using square root sequence points when varying K , the number of square root sequence points (standard errors in parentheses) (random effect-multivariate t_3).

K	β_0	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3	\hat{I}_{max}	iteration
10,000	0.98 (0.50)	-2.92 (0.65)	-0.94 (0.23)	3.69 (0.31)	0.00042 (0.0043)	0.00035 (0.0018)	0.0061 (0.0048)	0.0024 (0.0040)	0.26 (0.0017)	-0.038 (0.043)	0.0084 (0.053)	-220.14	9
20,000	0.72 (0.42)	-2.93 (0.60)	-0.89 (0.36)	3.72 (0.59)	0.00069 (0.0029)	0.0028 (0.0094)	-0.0078 (0.0028)	0.0054 (0.0058)	0.18 (0.0042)	0.0019 (0.072)	0.0089 (0.043)	-219.48	3
30,000	1.23 (0.37)	-3.03 (0.48)	-0.82 (0.29)	3.84 (0.62)	0.000017 (0.00038)	0.0059 (0.0028)	-0.0028 (0.0000)	-0.0026 (0.0029)	0.28 (0.0000)	0.0046 (0.031)	0.00029 (0.036)	-218.29	5
40,000	1.26 (0.32)	-2.98 (0.50)	-0.94 (0.22)	3.69 (0.53)	0.00043 (0.0000)	0.00060 (0.00016)	-0.0035 (0.00053)	0.0043 (0.0013)	0.39 (0.0025)	0.0042 (0.0028)	0.0013 (0.034)	-219.63	2
50,000	1.28 (0.30)	-3.02 (0.44)	-0.92 (0.25)	3.73 (0.67)	-0.00053 (0.0000)	0.00028 (0.00019)	0.00073 (0.0000)	-0.00023 (0.0000)	0.42 (0.0038)	-0.0028 (0.045)	0.0046 (0.027)	-218.94	3
60,000	1.23 (0.30)	-2.99 (0.46)	-0.91 (0.24)	3.69 (0.68)	0.00043 (0.00035)	0.00039 (0.0000)	-0.0072 (0.0022)	0.00034 (0.0000)	0.43 (0.0027)	-0.0028 (0.048)	0.0028 (0.0075)	-216.42	2
70,000	1.27 (0.34)	-3.04 (0.44)	-0.92 (0.10)	3.73 (0.64)	0.00054 (0.00048)	0.00047 (0.0000)	0.0082 (0.0035)	0.0041 (0.0027)	0.41 (0.081)	-0.0039 (0.023)	0.00046 (0.037)	-216.21	2
80,000	1.28 (0.39)	-3.02 (0.52)	-0.94 (0.19)	3.76 (0.65)	-0.00037 (0.0025)	-0.0027 (0.0000)	0.0083 (0.0028)	0.0029 (0.0027)	0.42 (0.034)	-0.00022 (0.052)	-0.00056 (0.030)	-216.65	2
90,000	1.21 (0.23)	-2.92 (0.43)	-0.88 (0.24)	3.77 (0.62)	-0.00029 (0.0073)	-0.0026 (0.0073)	0.0078 (0.0076)	-0.0036 (0.00058)	0.37 (0.00028)	0.0064 (0.055)	-0.0023 (0.029)	-216.96	1
100,000	1.29 (0.40)	-3.05 (0.48)	-0.93 (0.17)	3.72 (0.55)	-0.00094 (0.0031)	-0.0018 (0.0026)	0.0089 (0.0065)	0.0032 (0.0089)	0.42 (0.0097)	-0.0029 (0.046)	-0.00041 (0.033)	-216.34	2

Table 2. MLEs of the parameters by covariance modelling for the pooled salamander data using square root sequence points when varying K , the number of square root sequence points (standard errors in parentheses) (random effect-multivariate t_7).

K	β_0	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3	\hat{I}_{max}	iteration
10,000	0.58 (0.36)	-2.68 (0.52)	-0.98 (0.43)	3.77 (0.67)	0.00056 (0.0026)	0.00023 (0.0052)	0.00089 (0.0059)	0.0061 (0.0049)	0.47 (0.0036)	0.0014 (0.023)	0.024 (0.031)	-212.37	7
20,000	0.64 (0.50)	-2.85 (0.49)	-0.89 (0.32)	3.61 (0.59)	0.00012 (0.00023)	0.0036 (0.0039)	0.00056 (0.00067)	0.0056 (0.0048)	0.49 (0.0002)	-0.0016 (0.033)	-0.0023 (0.020)	-209.58	4
30,000	1.34 (0.33)	-2.98 (0.42)	-0.97 (0.35)	3.72 (0.64)	0.00023 (0.0000)	0.00053 (0.0083)	-0.0036 (0.0032)	-0.0032 (0.0037)	0.30 (0.0000)	0.0024 (0.042)	0.00042 (0.023)	-210.32	4
40,000	1.26 (0.38)	-3.05 (0.35)	-0.95 (0.24)	3.66 (0.53)	0.00017 (0.00024)	0.00024 (0.00037)	-0.0082 (0.00038)	0.0021 (0.0016)	0.39 (0.0000)	0.0032 (0.0034)	0.0034 (0.0037)	-211.26	3
50,000	1.28 (0.41)	-3.05 (0.37)	-0.96 (0.27)	3.70 (0.64)	0.00037 (0.0000)	0.00054 (0.00015)	0.00027 (0.0000)	0.00072 (0.0012)	0.43 (0.0027)	-0.0013 (0.013)	0.0027 (0.0092)	-207.93	2
60,000	1.27 (0.41)	-3.11 (0.42)	-0.99 (0.21)	3.68 (0.59)	0.00043 (0.0000)	0.00082 (0.00034)	-0.0034 (0.0032)	0.00083 (0.0000)	0.46 (0.0039)	-0.0035 (0.034)	0.0047 (0.0023)	-210.29	3
70,000	1.29 (0.38)	-3.11 (0.38)	-0.96 (0.20)	3.69 (0.56)	0.00072 (0.00023)	0.00036 (0.0000)	0.0055 (0.0048)	0.00023 (0.0000)	0.49 (0.032)	-0.0032 (0.018)	0.00039 (0.0034)	-210.34	2
80,000	1.28 (0.41)	-3.08 (0.39)	-0.98 (0.24)	3.66 (0.54)	-0.00031 (0.00031)	-0.0052 (0.00027)	0.0024 (0.0023)	0.0042 (0.0000)	0.47 (0.025)	-0.0037 (0.039)	-0.00027 (0.0090)	-210.55	1
90,000	1.23 (0.35)	-3.06 (0.52)	-0.89 (0.19)	3.70 (0.36)	-0.00079 (0.00015)	-0.0036 (0.0014)	0.0030 (0.0045)	-0.0013 (0.00032)	0.37 (0.00082)	0.0029 (0.021)	-0.0016 (0.0024)	-210.16	2
100,000	1.27 (0.43)	-3.10 (0.45)	-0.97 (0.20)	3.68 (0.55)	-0.00034 (0.00028)	-0.0019 (0.0035)	0.0043 (0.0023)	0.0012 (0.0013)	0.45 (0.0037)	-0.0036 (0.035)	-0.00034 (0.0027)	-210.38	2

Table 3. MLEs of the parameters by covariance modelling for the pooled salamander data using square root sequence points when varying K , the number of square root sequence points (standard errors in parentheses) (random effect-multivariate t_{10}).

K	β_0	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3	\hat{I}_{max}	iteration
10,000	0.72 (0.49)	-2.99 (0.78)	-0.82 (0.39)	3.66 (0.07)	0.0003 (0.0042)	0.0004 (0.0037)	0.0036 (0.0034)	0.0020 (0.0029)	0.24 (0.0051)	-0.046 (0.056)	-0.0098 (0.046)	-212.32	2
20,000	0.68 (0.34)	-2.93 (0.61)	-0.68 (0.67)	3.61 (0.61)	0.00082 (0.0072)	0.0039 (0.0087)	-0.0098 (0.0078)	0.0047 (0.0070)	0.098 (0.044)	0.0093 (0.067)	0.0085 (0.034)	-209.67	3
30,000	1.27 (0.23)	-2.89 (0.53)	-1.02 (0.27)	3.68 (0.57)	-0.000017 (0.00082)	0.0027 (0.0092)	-0.0023 (0.0000)	-0.0042 (0.0000)	0.47 (0.0000)	0.0064 (0.024)	0.00039 (0.027)	-211.24	4
40,000	1.24 (0.31)	-2.88 (0.58)	-0.93 (0.33)	3.68 (0.51)	-0.0004 (0.00094)	0.00063 (0.00067)	-0.0013 (0.00036)	0.0029 (0.00083)	0.39 (0.0036)	0.0041 (0.0065)	0.0061 (0.035)	-212.36	2
50,000	1.23 (0.30)	-2.87 (0.52)	-0.92 (0.29)	3.72 (0.59)	0.00025 (0.00067)	0.00087 (0.0000)	0.000023 (0.0000)	-0.00021 (0.0000)	0.44 (0.0000)	0.00026 (0.034)	0.0052 (0.018)	-211.04	2
60,000	1.20 (0.32)	-2.85 (0.49)	-0.97 (0.31)	3.69 (0.63)	0.00019 (0.0000)	0.00093 (0.0000)	-0.000026 (0.0000)	0.00029 (0.0000)	0.48 (0.0024)	-0.00038 (0.062)	0.0038 (0.0087)	-211.32	2
70,000	1.20 (0.33)	-2.86 (0.51)	-0.98 (0.092)	3.72 (0.62)	0.00032 (0.00031)	0.00072 (0.0000)	0.0015 (0.00035)	-0.0031 (0.0000)	0.51 (0.052)	-0.00022 (0.035)	0.0063 (0.043)	-211.34	3
80,000	1.22 (0.40)	-2.97 (0.52)	-0.94 (0.25)	3.75 (0.65)	-0.00051 (0.0024)	-0.0023 (0.00071)	0.0046 (0.0000)	-0.0025 (0.0034)	0.56 (0.028)	0.00097 (0.027)	-0.00026 (0.031)	-211.89	1
90,000	1.19 (0.28)	-2.96 (0.47)	-0.89 (0.24)	3.73 (0.63)	-0.00062 (0.0026)	-0.0042 (0.0075)	0.0031 (0.0083)	-0.0076 (0.00045)	0.51 (0.00039)	-0.0028 (0.052)	-0.0033 (0.025)	-211.83	2
100,000	1.27 (0.40)	-3.04 (0.52)	-0.90 (0.23)	3.70 (0.65)	-0.00071 (0.0031)	-0.0034 (0.0023)	0.0051 (0.00054)	-0.0024 (0.0089)	0.46 (0.077)	-0.0025 (0.055)	-0.00034 (0.023)	-211.52	3

Table 4. MLEs of the parameters by covariance modelling for the pooled salamander data using square root sequence points when varying K , the number of square root sequence points (standard errors in parentheses) (random effect-multivariate t_{15}).

K	β_0	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3	\hat{I}_{max}	iteration
10,000	0.92 (0.37)	-2.99 (0.34)	-0.57 (0.31)	3.72 (0.07)	0.00060 (0.0033)	0.00071 (0.0069)	0.0058 (0.0077)	0.0046 (0.0069)	0.28 (0.052)	-0.037 (0.058)	-0.032 (0.037)	-212.53	5
20,000	0.59 (0.36)	-2.90 (0.49)	-0.54 (0.55)	3.64 (0.64)	0.0013 (0.0030)	0.0012 (0.0062)	0.0062 (0.0082)	0.0052 (0.023)	0.087 (0.064)	0.0079 (0.059)	-0.0004 (0.052)	-210.62	3
30,000	1.28 (0.15)	-2.92 (0.35)	-1.00 (0.23)	3.75 (0.52)	0.000031 (0.0006)	0.0027 (0.0002)	-0.0020 (0.0000)	0.0033 (0.0000)	0.45 (0.0003)	0.0043 (0.025)	0.00063 (0.011)	-209.04	4
40,000	1.22 (0.33)	-2.92 (0.50)	-0.97 (0.24)	3.68 (0.57)	-0.00046 (0.0001)	0.00089 (0.0000)	-0.0043 (0.0000)	0.0026 (0.0000)	0.39 (0.024)	0.0021 (0.043)	0.0020 (0.029)	-209.01	2
50,000	1.24 (0.29)	-2.93 (0.48)	-0.96 (0.29)	3.67 (0.54)	0.00033 (0.0000)	0.00088 (0.0000)	0.000059 (0.0000)	0.0014 (0.0001)	0.45 (0.0000)	0.00020 (0.032)	0.0045 (0.020)	-209.21	2
60,000	1.20 (0.30)	-2.89 (0.51)	-0.96 (0.26)	3.74 (0.59)	0.00042 (0.0000)	0.00083 (0.0003)	-0.00008 (0.0000)	0.00081 (0.0000)	0.41 (0.024)	-0.00038 (0.037)	0.0019 (0.026)	-209.39	1
70,000	1.23 (0.35)	-2.86 (0.49)	-0.93 (0.070)	3.70 (0.53)	0.00054 (0.0000)	0.00079 (0.0000)	-0.00026 (0.0001)	-0.0051 (0.0000)	0.56 (0.064)	-0.00044 (0.063)	0.0020 (0.050)	-209.98	2
80,000	1.23 (0.37)	-2.92 (0.45)	-0.92 (0.20)	3.75 (0.61)	-0.00092 (0.0012)	-0.0012 (0.0022)	0.0068 (0.0004)	-0.0035 (0.0047)	0.50 (0.0013)	0.0021 (0.044)	0.00013 (0.027)	-209.13	2
90,000	1.18 (0.29)	-2.87 (0.42)	-0.87 (0.24)	3.72 (0.59)	-0.00048 (0.0020)	-0.0045 (0.0077)	0.0054 (0.0055)	-0.0078 (0.0000)	0.47 (0.0000)	-0.0042 (0.031)	-0.0024 (0.023)	-209.48	3
100,000	1.27 (0.40)	-2.99 (0.50)	-0.91 (0.22)	3.69 (0.56)	-0.00082 (0.0022)	-0.0030 (0.0040)	0.0062 (0.0000)	-0.0022 (0.0034)	0.43 (0.021)	-0.0022 (0.026)	-0.00036 (0.025)	-207.86	3

20-dimensional unit cube so that even if the number of the QMC points is chosen as $K = 100000$, it is still small in such a space. Nevertheless, the numerical results show that the QMC points approximation based estimation in the GLMMs performs very well in terms of accuracy and stabilization. When using 100,000 QMC points, our Fortran code takes about 50 minutes to obtain the results. Furthermore, the convergence of our algorithm is usually made between the 3th and 5th iterations. On the other hand, increasing the number of points may lead to less iterations.

For mean model, the estimators for fixed effects, β are similar when K increases over 30,000. For autoregressive coefficients (ACs) model, the estimations of γ , including $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$ and $\hat{\gamma}_4$ are very close to zero. For logarithms of innovation variances (IVs), only $\hat{\gamma}_1$ is not close to zero. It needs to be noted that some of standard deviations are 0.0000 for $\hat{\gamma}$ and $\hat{\lambda}$, which means those values are less than 10^{-7} .

A criteria easy to be executed to decide which covariate is significant is if the estimator is greater than twice of standard error. So the four fixed effect parameters of $\hat{\beta}$ are significant; four parameters for $\hat{\gamma}$ are not; $\hat{\lambda}_1$ is significant, but $\hat{\lambda}_2$ and $\hat{\lambda}_3$ are not. The conclusion is that $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4, \hat{\lambda}_1$ and $\hat{\lambda}_3$ can be regarded as zero. In other words, T is an identity matrix and D is a diagonal matrix with the same element. So $\hat{\Sigma}_b = \hat{\sigma}^2 I_q$, it can be said that every dimension of random effect of each salamander is independent of others. Also, the effects of female and male salamanders are independent and no heterogeneity between female and male.

Another point is why the random effect is not assumed as multivariate mixture normal distribution for the salamander data because Newton-Raphson algorithm does not achieve convergency until the mixture normal distribution degenerates to normal distribution.

6. Simulation

We conducted a simulation study to assess the efficiency of the QMC estimation in the GLMMs, in particular, to assess the performance of the GP set point when the distribution of random effect is multivariate t distribution (1) and multivariate mixture normal distribution (2). Based on the logistic model in (13), we simulate the salamander pooling data, which have 360 correlated binary observations. In the simulation study, we run 100 simulations. The same protocol design as the salamander mating experiment is adopted for the simulated data and the log-likelihood function for each simulated data set thus involves six 20-dimensional integrals that are analytically intractable.

When using the QMC approximation, we generate $K = 50000$ integration nodes on the cube C^{20} using the square roots of the first 20 prime numbers (the GP set method). Then, we use Equation (6) to approximate the integrated log-likelihood when random effect followed multivariate t distribution. When random effect followed multivariate mixture normal distribution, the log-likelihood changes to

$$\begin{aligned} \ell(\beta, \theta) &= \log L(\beta, \theta) = \log \left(\frac{1}{K} \sum_{k=1}^K \exp \left\{ \sum_{i=1}^n \ell_i(\beta, \theta | b_k) \right\} \right) \\ &= \log \left(\frac{1}{K} \sum_{k=1}^K \exp \left\{ \sum_{i=1}^n \ell_i \left(\beta, \theta \mid \left[\Sigma^{\frac{1}{2}}(\theta) \Phi^{-1}(c_k) + \mu_1 \right] I(U \leq \pi) + \left[\Sigma^{\frac{1}{2}}(\theta) \Phi^{-1}(c_k) + \mu_2 \right] I(U > \pi) \right) \right\} \right) \end{aligned} \quad (15)$$

The Newton-Raphson algorithm (10) is used to maximize the approximated log-likelihood function. The true values are chosen to be (1.20, -2.80, -1.00, 3.60) for the fixed effects, (-1.2, 0.6, -0.1, 0.4) for ACs parameters and (-2, 1, 0.5) for IVs parameters. ACs parameters and IVs parameters build variance components. The starting values for the algorithm are chosen to be the estimators as for the real data analysis. The convergence criterion is set to be 10^{-2} for the successive iterated values. The algorithm stops after 20 iterations and is considered to be non-convergent. The simulation studies are conducted and coded in FORTRAN language and this programme was run on a PC Pentium (R) 4 PC (CPU 3.20 GHz). A summary of the simulation results is presented in **Table 5** for multivariate t distribution and **Table 6** for multivariate mixture normal distribution. The results given are the average parameter estimates based on 100 replications and their standard errors.

It is clear that estimation by modified Cholesky decomposition with QMC approach is able to produce reasonably good results in GLMMs when the random effect is not normal distribution. When the random effects followed by multivariate t distribution (**Table 5**), the estimates of regression fixed effects and variance components are those on average of 100 simulation, which has little bias. That means the method of modified

Table 5. The average of the 100 parameter estimates in the simulation study, where the QMC approximation uses $K = 50000$ points implemented using the gp set for random effect of multivariate t_3 , t_7 , t_{10} and t_{15} distribution (simulated standard errors are given in parentheses).

	β_0	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4
True	1.20	-2.80	-1.00	3.60	-1.20	0.60	-0.10	0.40
t_3	1.22	-2.95	-0.98	3.72	-1.20	0.60	-0.10	0.40
StD	0.055	0.101	0.0950	0.179	-0.0000	-0.0003	0.0000	0.0000
t_7	1.18	-2.75	-0.88	3.51	-1.22	0.59	-0.11	0.43
StD	0.05	0.10	0.10	0.19	-0.0001	0.0019	0.0053	0.0000
t_{10}	1.22	-2.86	-1.07	3.80	-1.20	0.61	-0.12	0.39
StD	0.043	0.065	0.083	0.12	0.0000	-0.0007	0.0004	-0.0042
t_{15}	1.19	-2.84	-0.99	3.64	-1.20	0.62	-0.086	0.39
StD	0.049	0.103	0.086	0.187	0.0001	-0.0067	0.0035	0.0072
	λ_1	λ_2	λ_3	$\hat{\ell}_{\max}$				
True	-2.00	-1.00	0.50					
t_3	-2.00	-1.00	0.50	-211.00				
StD	-0.0000	-0.0000	-0.0000					
t_7	-1.99	-1.02	0.52	-210.47				
StD	-0.0001	0.0002	-0.0001					
t_{10}	-1.98	-0.97	0.52	-211.36				
StD	-0.0000	0.0014	0.0001					
t_{15}	-2.01	-0.99	0.51	-209.93				
StD	0.0024	0.0024	0.0026					

Table 6. The average of the 100 parameter estimates in the simulation study, where the QMC approximation uses $K = 50000$ points implemented using the gp set for random effect of multivariate mixture normal distribution by $\pi = 0.25$, $\pi = 0.5$ and $\pi = 0.75$ (simulated standard errors are given in parentheses).

	β_0	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4
True	1.20	-2.80	-1.00	3.60	-1.20	0.60	-0.10	0.40
$\pi = 0.25$	1.29	-2.84	-0.97	3.45	-1.20	0.67	-0.085	0.43
StD	0.050	0.11	0.10	0.19	0.0001	-0.0012	-0.0025	0.0002
$\pi = 0.5$	1.36	-3.20	-0.93	3.70	-1.21	0.57	-0.083	0.40
StD	0.044	0.075	0.097	0.13	0.0000	0.0002	-0.0012	0.0026
$\pi = 0.75$	1.39	-3.16	-1.12	3.94	-1.20	0.64	-0.062	0.36
StD	0.049	0.077	0.094	0.14	-0.0000	-0.056	-0.0026	-0.0060
	λ_1	λ_2	λ_3	$\hat{\ell}_{\max}$				
True	-2.00	-1.00	0.50					
$\pi = 0.25$	-2.00	-1.00	0.51	-212.10				
StD	0.0002	0.0016	0.0008					
$\pi = 0.5$	-2.06	-1.06	0.47	-226.41				
StD	0.0000	0.0007	0.0005					
$\pi = 0.75$	-1.99	-1.03	0.47	-299.15				
StD	0.0006	-0.0026	0.0002					

Cholesky decomposition of covariance modelling with QMC approach is quite useful and effective. When the random effects followed by multivariate mixture normal distribution (Table 6), the estimates of regression fixed effects are acceptable although bigger bias, especially for fixed effects. The estimates of variance components are quite accurate. That is, when random effects change to two peaks distribution, the accuracy need to be enhanced.

7. Conclusion and Discussion

It is quite common in the literatures that random effects are independent identically distributed normal variables in generalized linear models (GLMMs). In longitudinal study, the covariance-variance matrix of random effects for individuals are usually assumed to be homogeneous. However, this assumption may not be valid in practice. In this article, the assumption extends to random effects follow multivariate t and mixture normal distribution. Approximation of *marginal quasi-likelihood* and parameter estimation in GLMMs is very difficult and challenging because they involve integral on random effects, especially for high-dimensional problems.

In this article, we have used Quasi-Monte Carlo (QMC) sequence approximation to calculate the maximum likelihood estimates and marginal log quasi-likelihood in GLMMs. The key idea of QMC sequences is to generate integration points which are scattered uniformly in the integration domain. The good point, one simple QMC point, is used through this article. Newton-Raphson (NR) algorithm is the iterative method to obtain optimum. QMC-NR method converges very quickly and performs very well in terms of accuracy and stabilization. In longitudinal studies, the covariance structure plays a crucial role in statistical inferences. The correct modelling of covariance structure improves the efficiencies of the mean parameters and provides much more reliable estimates (Ye and Pan, 2006, [16]). The joint modelling (JM) for mean and for covariance-variance components for random effects reduces the number of parameters and has a clear statistical meaning. QMC-NR method with joint modelling can be called as QMC-NR-JM method. Although QMC-NR-JM method performs well in stability and produces accurate estimation of the parameters in GLMMs, there is no way to avoid intensive computation in real data analysis and simulation study. Particularly, the binary cross-designed salamander dataset, involving six 20-dimensional integrals in the log quasi-likelihood function, has been analyzed and the same protocol design is adopted for simulations.

Note that a practical issue is how to select of the number of integration points? My solution is increasing the number of QMC points gradually until the MLEs become stable. It correspondingly leads to increasing the computational time and efforts. This problem is quite obvious for simulation study, especially when the number of parameters increases. Another worth noting issue is that a range of starting values for the fixed effects and variance components have been tried and we find that the different starting values almost can't change the results when algorithm reaches convergency.

In the QMC-NR-JM approximation method, the estimators of random effects cannot be calculated at the same time when estimating MLEs for fixed effects and variance components. Pan and Thompson (2007) [7] proposed an iterative method to predict of the random effects, implying that an initial value of b substituted into the equation to yield a new prediction of b . This process is then repeated until the convergence of \hat{b} , but this method cannot guarantee convergence for arbitrary data. So how to estimate the random effects is still a open question in QMC-NR-JM method.

Acknowledgements

We thank the editors and the reviewers for their comments. This research is funded by the National Social Science Foundation No. 12CGL077, National Science Foundation granted No. 71201029, No.71303045 and No. 11561071. This support is greatly appreciated.

References

- [1] Breslow, N.E. and Clayton, D.G. (1993) Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **88**, 9-25.
- [2] Lin, X. and Breslow, N.E. (1996) Bias Correction in Generalized Linear Mixed Models with Multiple Components of Dispersion. *Journal of the American Statistical Association*, **91**, 1007-1016.
<http://dx.doi.org/10.1080/01621459.1996.10476971>

- [3] Lee, Y. and Nelder, J.A. (2001) Hierarchical Generalized Linear Models: A Synthesis of Generalized Linear Models, Random Effect Models and Structured Dispersions. *Biometrika*, **88**, 987-1006. <http://dx.doi.org/10.1093/biomet/88.4.987>
- [4] Karim, M.R. and Zeger, S.L. (1992) Generalized Linear Models with Random Effects; Salamnder Mating Revisited. *Biometrics*, **48**, 681-694. <http://dx.doi.org/10.2307/2532317>
- [5] Booth, J.G. and Hobert, J.P. (1999) Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm. *Journal of the Royal Statistical Society*, **61**, 265-285. <http://dx.doi.org/10.1111/1467-9868.00176>
- [6] Pan, J. and Thompson, R. (2003) Gauss-Hermite Quadrature Approximation for Estimation in Generalized Linear Mixed Models. *Computational Statistics*, **18**, 57-78.
- [7] Pan, J. and Thompson, R. (2007) Quasi-Monte Carlo Estimation in Generalized Linear Mixed Models. *Computational Statistics and Data Analysis*, **51**, 5765-5775. <http://dx.doi.org/10.1016/j.csda.2006.10.003>
- [8] Al-Eleid, E.M.O. (2007) Parameter Estimation in Generalized Linear Mixed Models Using Quasi-Monte Carlo Methods. PhD Dissertation, The University of Manchester, Manchester.
- [9] McCullouch, C.E. (2003) Generalized Linear Mixed Models. NSF-CBMS Regional Conference Series in Probability and Statistics Volume 7. Institution of Mathematical Statistics and the American Statistical Association, San Francisco.
- [10] McCulloch, P. and Nelder, J.A. (1989) Generalized Linear Models. 2nd Edition, Chapman and Hall, London. <http://dx.doi.org/10.1007/978-1-4899-3242-6>
- [11] Niederreiter, H. (1992) Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia. <http://dx.doi.org/10.1137/1.9781611970081>
- [12] Fang, K.T. and Wang, Y. (1994) Number-Theoretic Methods in Statistics. Chapman and Hall, London. <http://dx.doi.org/10.1007/978-1-4899-3095-8>
- [13] Pourahmadi, M. (1999) Joint Mean-Covariance Models with Applications to Longitudinal Data: Unconstrained Parameterization. *Biometrika*, **86**, 677-690. <http://dx.doi.org/10.1093/biomet/86.3.677>
- [14] Pourahmadi, M. (2000) Maximum Likelihood Estimation of Generalized Linear Models for Multivariate Normal Covariance Matrix. *Biometrika*, **87**, 425-435. <http://dx.doi.org/10.1093/biomet/87.2.425>
- [15] Shun, Z. (1997) Another Look at the Salamander Mating Data: A Modified Laplace Approximation Approach. *Journal of the American Statistical Association*, **92**, 341-349. <http://dx.doi.org/10.1080/01621459.1997.10473632>
- [16] Ye, H. and Pan, J. (2006) Modelling Covariance Structures in Generalized Estimating Equations for Longitudinal Data. *Biometrika*, **93**, 927-941. <http://dx.doi.org/10.1093/biomet/93.4.927>

Appendix A. Second-Order Derivatives of Log-Likelihood Newton-Raphson Algorithm

$$\begin{pmatrix} \beta \\ \theta \end{pmatrix}_{(m+1)} = \begin{pmatrix} \beta \\ \theta \end{pmatrix}_{(m)} + \begin{pmatrix} \frac{\partial^2 \ell}{\partial \beta \partial \beta'} & \frac{\partial^2 \ell}{\partial \beta \partial \theta'} \\ \frac{\partial^2 \ell}{\partial \theta \partial \beta'} & \frac{\partial^2 \ell}{\partial \theta \partial \theta'} \end{pmatrix}_{(m)}^{-1} \begin{pmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial \theta} \end{pmatrix}_{(m)}$$

where $\frac{\partial \ell}{\partial \beta}$ and $\frac{\partial \ell}{\partial \theta}$ are score function of log likelihood, $\begin{pmatrix} \frac{\partial^2 \ell}{\partial \beta \partial \beta'} & \frac{\partial^2 \ell}{\partial \beta \partial \theta'} \\ \frac{\partial^2 \ell}{\partial \theta \partial \beta'} & \frac{\partial^2 \ell}{\partial \theta \partial \theta'} \end{pmatrix}$ is Hessian matrix of log likelihood.

The second-order derivatives of log likelihood

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial \beta^T} &= \sum_{k=1}^K w_k \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} x_i \right] \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} x_i^T \right] \\ &\quad - \left[\sum_{k=1}^K w_k \sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} x_i \right] \left[\sum_{k=1}^K w_k \sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} x_i^T \right] \\ &\quad - \sum_{k=1}^K w_k \left(\sum_{i=1}^n \frac{a_i}{\phi} \left[v(\mu_{ik}^{b_k}) \dot{g}^2(\mu_{ik}^{b_k}) \right]^{-1} x_i x_i^T \right) \\ &\quad - \sum_{k=1}^K w_k \left(\sum_{i=1}^n \frac{a_i}{\phi} \left[\frac{\dot{v}(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k}) + v(\mu_{ik}^{b_k}) \ddot{g}(\mu_{ik}^{b_k})}{v^2(\mu_{ik}^{b_k}) \dot{g}^3(\mu_{ik}^{b_k})} \right] (y_i - \mu_{ik}^{b_k}) x_i x_i^T \right) \\ \frac{\partial^2 l}{\partial \theta_j \partial \theta_j} &= \sum_{k=1}^K w_k \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right] \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right] \\ &\quad - \left[\sum_{k=1}^K w_k \sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right]^2 - \sum_{k=1}^K w_k \left(\sum_{i=1}^n \frac{a_i}{\phi} \left[v(\mu_{ik}^{b_k}) \dot{g}^2(\mu_{ik}^{b_k}) \right]^{-1} \right. \\ &\quad \left. + \left[\frac{\dot{v}(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k}) + v(\mu_{ik}^{b_k}) \ddot{g}(\mu_{ik}^{b_k})}{v^2(\mu_{ik}^{b_k}) \dot{g}^3(\mu_{ik}^{b_k})} \right] \right) \left(z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right)^2 + \sum_{k=1}^K w_k \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right] \\ \frac{\partial^2 l}{\partial \theta_j \partial \theta_k} &= \sum_{k=1}^K w_k \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right] \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_k}^{\frac{1}{2}} F^{-1}(c_k) \right] \\ &\quad - \left[\sum_{k=1}^K w_k \sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right] \left[\sum_{k=1}^K w_k \sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_k}^{\frac{1}{2}} F^{-1}(c_k) \right] \\ &\quad - \sum_{k=1}^K w_k \left(\sum_{i=1}^n \frac{a_i}{\phi} \left[v(\mu_{ik}^{b_k}) \dot{g}^2(\mu_{ik}^{b_k}) \right]^{-1} + \left[\frac{\dot{v}(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k}) + v(\mu_{ik}^{b_k}) \ddot{g}(\mu_{ik}^{b_k})}{v^2(\mu_{ik}^{b_k}) \dot{g}^3(\mu_{ik}^{b_k})} \right] \right) \left(z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right)^2 \\ &\quad + \sum_{k=1}^K w_k \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j, \theta_k}^{\frac{1}{2}} F^{-1}(c_k) \right] \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial \theta_j} &= \sum_{k=1}^K w_k \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right] \left[\sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} x_i \right] \\ &\quad - \left[\sum_{k=1}^K w_k \sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \right] \left[\sum_{k=1}^K w_k \sum_{i=1}^n \frac{a_i (y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k})} x_i \right] \\ &\quad - \sum_{k=1}^K w_k \left(\sum_{i=1}^n \frac{a_i}{\phi} \left[v(\mu_{ik}^{b_k}) \dot{g}^2(\mu_{ik}^{b_k}) \right]^{-1} z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \cdot x_i \right) \\ &\quad - \sum_{k=1}^K w_k \left(\sum_{i=1}^n \frac{a_i}{\phi} \left[\frac{\dot{v}(\mu_{ik}^{b_k}) \dot{g}(\mu_{ik}^{b_k}) + v(\mu_{ik}^{b_k}) \ddot{g}(\mu_{ik}^{b_k})}{v^2(\mu_{ik}^{b_k}) \dot{g}^3(\mu_{ik}^{b_k})} \right] (y_i - \mu_{ik}^{b_k}) z_i^T \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k) \cdot x_i \right) \end{aligned}$$

Appendix B: MLEs for Covariance Modeling

The first and second derivative of Σ_b respect to γ and λ are:

$$\dot{\Sigma}_{\gamma_j}^{\frac{1}{2}} = \frac{\partial \Sigma_b^{\frac{1}{2}}}{\partial \gamma_j} = \frac{\partial T^{-1}(\gamma)}{\partial \gamma_j} \cdot D^{\frac{1}{2}}(\lambda) \quad (j=1, \dots, q_1^*) \tag{16}$$

$$\ddot{\Sigma}_{\gamma_j}^{\frac{1}{2}} = \frac{\partial^2 \Sigma_b^{\frac{1}{2}}}{\partial \gamma_j^2} = \frac{\partial^2 T^{-1}(\gamma)}{\partial \gamma_j^2} \cdot D^{\frac{1}{2}}(\lambda) \quad (j=1, \dots, q_1^*) \tag{17}$$

$$\ddot{\Sigma}_{\gamma_j \gamma_k}^{\frac{1}{2}} = \frac{\partial^2 \Sigma_b^{\frac{1}{2}}}{\partial \gamma_j \partial \gamma_k} = \frac{\partial^2 T^{-1}(\gamma)}{\partial \gamma_j \partial \gamma_k} \cdot D^{\frac{1}{2}}(\lambda) \quad (j \neq k \text{ and } 1 \leq j, k \leq q_1^*) \tag{18}$$

Φ^* is introduced for the purpose of calculation for $\frac{\partial T^{-1}(\gamma)}{\partial \gamma_j}$, $\frac{\partial^2 T^{-1}(\gamma)}{\partial \gamma_j^2}$ and $\frac{\partial^2 T^{-1}(\gamma)}{\partial \gamma_j \partial \gamma_k}$. Note that the matrix Φ^* looks similar to the matrix T , but opposite sign for the lower triangular elements.

$$\Phi^* = \begin{pmatrix} 1 & & & & & \\ \phi_{21} & 1 & & & & \\ \phi_{31} & \phi_{32} & 1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ \vdots & \vdots & \vdots & & \ddots & \\ \phi_{q,1} & \phi_{q,2} & \phi_{q,3} & \cdots & \phi_{q,q-1} & 1 \end{pmatrix} \tag{19}$$

Every element in Φ^* was rearranged into a response vector Φ , such that $\Phi = C\gamma$, i.e.,

$$\Phi = \begin{pmatrix} \phi_{21} \\ \phi_{31} \\ \phi_{32} \\ \vdots \\ \phi_{q,q-1} \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_i \\ \vdots \\ \frac{\phi_{q(q-1)}}{2} \end{pmatrix} = \begin{pmatrix} c_{1,1} & \cdots & c_{1,j} & \cdots & c_{1,q_1^*} \\ \vdots & & \vdots & & \vdots \\ c_{i,1} & \cdots & c_{i,j} & \cdots & c_{i,q_1^*} \\ \vdots & & \vdots & & \vdots \\ c_{\frac{q(q-1)}{2},1} & \cdots & c_{\frac{q(q-1)}{2},j} & \cdots & c_{\frac{q(q-1)}{2},q_1^*} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_j \\ \vdots \\ \gamma_{q_1^*} \end{pmatrix} \tag{20}$$

We can put the j^{th} column of C into a $q \times q$ lower triangular matrix like Φ^* . So we have $q_1^* q \times q$ matrices. The element $c_{i,j}$ in (20) can be put into the place (d_1, d_2) in the j^{th} matrix. For the j^{th} matrix, $j=1, \dots, q_1^*$, we have the relationship $i=0+1+2+\dots+(d_1-2)+d_2$ where $1 \leq d_1 \leq q$, $1 \leq d_2 \leq d_1-1$ and $1 \leq i \leq \frac{q(q-1)}{2}$.

The above relationship is a one-to-one correspondence between i and (d_1, d_2) . So ϕ_i and $c_{i,j}$ in (20) can be also expressed as

$$\phi_i \Leftrightarrow \phi_{d_1, d_2} \quad c_{i,j} \Leftrightarrow c_{d_1, d_2}^j \tag{21}$$

T^{-1} is a lower triangular matrix with 1's as diagonal entries, so that $\frac{\partial T^{-1}(\gamma)}{\partial \gamma_j}$, $\frac{\partial^2 T^{-1}(\gamma)}{\partial \gamma_j^2}$ and $\frac{\partial^2 T^{-1}(\gamma)}{\partial \gamma_j \partial \gamma_k}$ are all lower triangular matrices. Denote $T^{-1} = (a_{d_1, d_2})$ where

$$a_{d_1, d_2} = \begin{cases} \phi_{d_1 d_2} & d_1 = d_2 + 1 \\ \phi_{d_1 d_2} + \sum_{m=d_2+1}^{d_1-1} \phi_{d_1 m} a_{m d_2} & d_1 \geq d_2 + 2 \end{cases} \tag{22}$$

The first derivative can be calculated as

$$\frac{\partial T^{-1}}{\partial \gamma_j} = -T^{-1} \cdot \frac{\partial T}{\partial \gamma_j} \cdot T^{-1} \tag{23}$$

where $\frac{\partial T}{\partial \gamma_j}$ is a lower triangular matrix with 0's as diagonal entries and $-c_{d_1, d_2}^j$ as the entries below diagonal.

The second derivative can be calculated as

$$\frac{\partial^2 T^{-1}}{\partial \gamma_j^2} = - \left[\frac{\partial T^{-1}}{\partial \gamma_j} \cdot \frac{\partial T}{\partial \gamma_j} \cdot T^{-1} + T^{-1} \cdot \frac{\partial T}{\partial \gamma_j} \cdot \frac{\partial T^{-1}}{\partial \gamma_j} \right] \tag{24}$$

$$\frac{\partial^2 T^{-1}}{\partial \gamma_j \partial \gamma_k} = - \left[\frac{\partial T^{-1}}{\partial \gamma_k} \cdot \frac{\partial T}{\partial \gamma_j} \cdot T^{-1} + T^{-1} \cdot \frac{\partial T}{\partial \gamma_j} \cdot \frac{\partial T^{-1}}{\partial \gamma_k} \right] \tag{25}$$

because of $\frac{\partial^2 T}{\partial \gamma_j^2} = 0$, $\frac{\partial^2 T}{\partial \gamma_j \partial \gamma_k} = 0$. We also have

$$\dot{\Sigma}_{\lambda_j}^{\frac{1}{2}} = \frac{\partial \Sigma_b^{\frac{1}{2}}}{\partial \lambda_j} = T^{-1}(\gamma) \cdot \frac{\partial D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j} \quad (j = 1, \dots, q_2^*) \tag{26}$$

$$\ddot{\Sigma}_{\lambda_j}^{\frac{1}{2}} = \frac{\partial^2 \Sigma_b^{\frac{1}{2}}}{\partial \lambda_j^2} = T^{-1}(\gamma) \cdot \frac{\partial^2 D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j^2} \quad (j = 1, \dots, q_2^*) \tag{27}$$

$$\ddot{\Sigma}_{\lambda_j \lambda_k}^{\frac{1}{2}} = \frac{\partial^2 \Sigma_b^{\frac{1}{2}}}{\partial \lambda_j \partial \lambda_k} = T^{-1}(\gamma) \cdot \frac{\partial^2 D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j \partial \lambda_k} \quad (j \neq k \text{ and } 1 \leq j, k \leq q_2^*) \tag{28}$$

where

$$D^{\frac{1}{2}}(\lambda) = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_q \end{pmatrix} = \begin{pmatrix} \exp\left(\frac{1}{2} h'_1 \lambda\right) & & \\ & \ddots & \\ & & \exp\left(\frac{1}{2} h'_q \lambda\right) \end{pmatrix} \tag{29}$$

$$\frac{\partial D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j} = \frac{1}{2} \begin{pmatrix} \sigma_1 h_{1j} & & \\ & \sigma_2 h_{2j} & \\ & & \ddots \\ & & & \sigma_q h_{q,j} \end{pmatrix} = \frac{1}{2} D^{\frac{1}{2}}(\lambda) \begin{pmatrix} h_{1j} & & \\ & h_{2j} & \\ & & \ddots \\ & & & h_{q,j} \end{pmatrix} \tag{30}$$

$$\frac{\partial^2 D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j^2} = \frac{1}{2} \frac{\partial D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j} \begin{pmatrix} h_{1j} & & & \\ & h_{2j} & & \\ & & \ddots & \\ & & & h_{q,j} \end{pmatrix} = \frac{1}{4} D^{\frac{1}{2}}(\lambda) \begin{pmatrix} h_{1j}^2 & & & \\ & h_{2j}^2 & & \\ & & \ddots & \\ & & & h_{q,j}^2 \end{pmatrix} \quad (31)$$

$$\frac{\partial^2 D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j \partial \lambda_k} = \frac{1}{2} \frac{\partial D^{\frac{1}{2}}(\lambda)}{\partial \lambda_k} \begin{pmatrix} h_{1j} & & & \\ & h_{2j} & & \\ & & \ddots & \\ & & & h_{q,j} \end{pmatrix} = \frac{1}{4} D^{\frac{1}{2}}(\lambda) \begin{pmatrix} h_{1j} h_{1k} & & & \\ & h_{2j} h_{2k} & & \\ & & \ddots & \\ & & & h_{q,j} h_{q,k} \end{pmatrix} \quad (32)$$