Quasi-Monte Carlo Estimation in Generalized Linear Mixed Model with Correlated Random Effects

Yin Chen[1], Yu Fei[2], Jianxin Pan[3]*

[1]School of Insurance, University of International Business and Economics, Beijing, China
[2]School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, China
[3]School of Mathematics, University of Manchester, Manchester, UK
Email: chenyin@uibe.edu.cn, feiyukm@aliyun.com, *jianxin.pan@manchester.ac.uk

Abstract

Parameter estimation by maximizing the marginal likelihood function in generalized linear mixed models (GLMMs) is highly challenging because it may involve analytically intractable high-dimensional integrals. In this paper, we propose to use Quasi-Monte Carlo (QMC) approximation through implementing Newton-Raphson algorithm to address the estimation issue in GLMMs. The random effects release to be correlated and joint mean-covariance modelling is considered. We demonstrate the usefulness of the proposed QMC-based method in approximating high-dimensional integrals and estimating the parameters in GLMMs through simulation studies. For illustration, the proposed method is used to analyze the infamous salamander mating binary data, of which the marginalized likelihood involves six 20-dimensional integrals that are analytically intractable, showing that it works well in practices.

Keywords

1. Introduction

McCullagh and Nelder (1989) [1] analyzed a challenging salamander mating binary data set which was conducted on two geographically isolated populations, Roughbutt (RB) and Whiteside (WS), in three experiments

---

*Corresponding author.

(one in summer 1986 and two in fall 1986). The question of interest is to determine whether or not certain barriers to interbreeding have evolved over time in geographically isolated populations. The second question is to determine whether or not heterogeneity between individuals (females and males) exists. The RB and WS populations are ideal for these experiments because salamanders would never meet those from different population in their natural environment, so that there are two closed groups in each experiment. Each group involved five females and five males from each population. Each female mated with six males, three from each population. So each closed group resulted in sixty correlated binary observations (1 for successful mating; 0 for failed mating). Therefore, the binary data set is crossed, 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). The data can be analyzed by using a specific generalized linear mixed model.

Generalized linear mixed models (GLMMs) are very useful for non-Gaussian correlated or clustered data and widely applied in many areas including epidemiology, ecological, and clinical trials. Parameters estimation in GLMMs is, however, very challenging, and correlated binary data is even more difficult to analyze than continuous data. This is because the marginalized likelihood function, obtained by integrating out multivariate random effects from the joint density function of the responses and random effects, is in general analytically intractable. For example, the likelihood function for the pooled salamander mating data described above involves six 20-dimensional integrals which are analytically intractable. In the literature, the marginalized likelihood function was considered by two typical methods: a) analytical Laplace approximation to integrals: This is represented by the work of Breslow and Clayton (1993) [2] who developed the penalized quasi-likelihood (PQL) estimation. Unfortunately, approaches based on Laplace approximation may yield biased estimates of variance components, particularly for modeling correlated binary data. This difficulty has stimulated a number of alternative approaches. For example, Lin and Breslow (1996) [3] proposed a corrected penalized quasi-likelihood, and Lee and Nelder (2001) [4] developed hierarchical generalized linear models (HGLMs) procedure; b) numerical techniques to approximate the likelihood function and estimate the parameters: Karim and Zeger (1992) [5] developed a Bayesian approach with Gibbs sampling for GLMMs, but its computational efforts may be computationally intensive. Monte Carlo Expectation-Maximization (MCEM) and Monte Carlo Newton-Raphson (MCNR) algorithms were proposed to approximate the maximum likelihood estimates in GLMMs by McCulloch (1994, 1997) [6] [7]. However, iterations of MCEM and MCNR do not always converge to the global maximum and are also very computationally intensives for high-dimensional problems (Chan, Kuk and Yam, 2005) [8]. Pan and Thompson (2003, 2007) [9] [10] developed Gauss-Hermite Quadrature (GHQ) approach and Quasi-Monte Carlo (QMC) approximation. Aleid and Pan (2008) [11] used Randomized Quasi-Monte Carlo (RQMC) to solve the integration problem in GLMMs. It is noted that most of the work above assume that random effects in GLMMs, which characterize the heterogeneity among individuals, are mutually independent and normally distributed.

However, these assumptions may not always hold due to the likely correlation of random effects, for instance. At least, the independence assumption of random effects should be testable. Misspecification of random effects covariance structure may yield inefficient estimates of parameters and can cause the problem of biased estimates of variance components, making statistical inferences for GLMMs unreliable. In this paper, we release the assumption of independence for random effects and aim to develop a new methodology which can model the covariance structures of random effects in GLMMs. Covariance modeling strategy based on a modified Cholesky decomposition is studied within the framework of GLMMs. The paper is organized as follows. GLMMs and the marginalized Quasi-likelihood are briefly reviewed in Section 2. The QMC approximation, the simplest good point set and analytical form of the MLEs for GLMMs are discussed in Section 3. Covariance modeling strategy through a modified Cholesky decomposition is described in Section 4. In Section 5, the salamander data set is analyzed as an example for illustration of the use of the new method. Simulation studies with the same design protocol as the salamander data are conducted in Section 6. Discussions on the related issues and further studies are presented in Section 7.

2. Generalized Linear Mixed Models

Let y_i $(i = 1, \cdots, n)$ denote the observed responses. Assume that x_i' is the ith row of design matrix X; β is the $p \times 1$ fixed effects parameter; z_i' is the ith row of the second design matrix Z; and b is the $q \times 1$ random effects.

The GLMMs are of the form

$$g\left(\mu_i^b\right) = \eta_i^b = x_i'\beta + z_i'b \,, \tag{1}$$

where $g(\cdot)$ is a monotonic and differentiable link function. Given random effects b, the responses y_i are independent and follow an exponential family of distributions with the density

$$f\left(y_i \middle| b\right) = \exp\left\{\frac{y_i\vartheta_i - b(\vartheta_i)}{a_i(\varphi)} + c(y_i, \varphi)\right\}. \tag{2}$$

The conditional mean and variances, given random effects b, are given by $\mu_i^b = E\left(Y_i \middle| b = b\right) = \dot{b}(\vartheta_i)$ and $var\left(Y_i \middle| b = b\right) = \frac{\varphi}{a_i}\ddot{b}(\vartheta_i) = \frac{\varphi}{a_i}v(\mu_i^b)$, respectively, where a_i is a prior weight, φ is an overdispersion scalar, ϑ_i is the so-called natural parameter, $\dot{b}(\vartheta_i)$ and $\ddot{b}(\vartheta_i)$ are the first- and second-derivatives of $b(\vartheta_i)$, and $v(\cdot)$ is the known variance function (McCullagh and Nelder, 1989) [1].

When the linear predictor η_i^b is identical to ϑ_i, i.e., $\eta_i^b = \vartheta_i$, GLMMs are said to have a canonical link function. In this article we assume the model has a canonical link function, although the proposed method can be extended straightforwardly to non-canonical link functions. Like generalized linear models (GLMs), the choice of canonical link function depends on the distribution of data. For example, for binary responses the canonical link function is the logit function, which takes the form

$$\text{logit}\left(\mu_i^b\right) = x_i'\beta + z_i'b = \eta_i^b = \log\left(\frac{\mu_i^b}{1 - \mu_i^b}\right). \tag{3}$$

Note q-dimensional random effects b are assumed to be normally distributed with mean 0 and positive definite variance-covariance matrix $\Sigma_b(\theta)$, where θ is an m-variant vector of variance components.

Quasi-likelihood function of the fixed effects β and variance components θ in GLMMs, also called *marginal quasi-likelihood function*, can be expressed by

$$L(\beta, \theta) = \exp\left\{\ell(\beta, \theta)\right\} = \int \exp\left\{\sum_{i=1}^n \ell_i(\beta, \theta \middle| b)\right\} dF(b; \theta), \tag{4}$$

where

$$\ell_i(\beta, \theta \middle| b) \propto \int_{y_i}^{\mu_i^b} \frac{a_i(y_i - u)}{\phi v(u)} du, \tag{5}$$

so that

$$\ell(\beta, \theta) = \log L(\beta, \theta) = \log \int \exp\left\{\sum_{i=1}^n \ell_i(\beta, \theta \middle| b)\right\} dF(b; \theta) \tag{6}$$

is the *conditional log quasi-likelihood* of β and θ, given random effects b (Breslow and Claybon, 1993) [2]. $F(b; \theta)$ is CDF of random effect b.

It is challenging to calculate the maximum likelihood estimates (MLEs), $\hat{\beta}$ and $\hat{\theta}$, that maximize the above conditional log quasi-likelihood, because it likely involves high-dimensional integrals which are analytically intractable. This computational problem becomes more severe when random effects are multivariate and correlated.

3. Quasi-Monte Carlo Integration and Estimation

3.1. Monte Carlo (MC) Integration

In this section, Quasi-Monte Carlo (QMC) approach is used to approximate high dimensional integrals. The traditional Monte Carlo (MC) method is outlined first as follow. Let $f(\cdot)$ be an integrable function over the q-dimensional unit cube $C^q = [0,1)^q$. Without loss of generality, we consider the following integrals

$$I(f) = \int_{C^q} f(x) dx = \int_0^1 \int_0^1 \cdots \int_0^1 f(x_1, x_2, \cdots, x_q) dx_1 dx_2 \cdots dx_q. \tag{7}$$

Assume K independent random samples $\{x_k : 1 \le k \le K\}$ are drawn from the uniform distribution on C^q and used to approximate the integral (7) through

$$\hat{I}_K(f, \mathcal{P}_K) = \frac{1}{K} \sum_{k=1}^{K} f(x_k). \tag{8}$$

According to the strong law of large number, the approximation $\hat{I}_K(f, \mathcal{P}_K)$ converges to the true value of the integral $I(f)$ with probability one as $K \to \infty$. The central limit theorem assures that the distribution of $\hat{I}_K(f, \mathcal{P}_K)$ converges to normal distribution with mean $I(f)$ and variance $\sigma_K^2 > 0$ as $K \to \infty$. The convergence rate for MC integration has an order $O(K^{-1/2})$, which is independent of the dimension of integral.

Since the rate of convergence is independent of q, the MC approximation seems to be very appealing. However, the MC algorithm has some serious drawbacks (see, e.g., Niederreiter, 1992 [12]; Spanier and Maize, 1994 [13]). One of these deficiencies is that even though the rate of convergence is independent of the dimension, the convergence is very slow. On the other hand, it may suffer from the problem of generating genuinely random samples. Instead, pseudo-random samples are generated and used to approximate the integral (Traub and Wozniakowski, 1992 [14]). It is also noted that the MC method only provides probabilistic error bounds, which is not a desirable as it does not guarantee reliable estimation results. As an alternative to the MC methods, Quasi-Monte Carlo approximation method is considered.

3.2. Quasi-Monte Carlo Method and Lattice Rule

Quasi-Monte Carlo (QMC) sequences are a deterministic alternative to Monte Carlo sequences (Niderreiter, 1992). In contrast to MC sequences, QMC sequences have a property that they are uniformly distributed on the unit cube. Thus, QMC approximation has the same equation as (8) but the MC random samples are replaced by deterministic samples that are uniformly distributed on the domain. Uniformity of sequences is measured by means of *discrepancy*, and for this reason QMC sequences are also called *low-discrepancy sequences*. Note the Koksma-Hlawka inequality

$$\left| \hat{I}_K(f, \mathcal{P}_K) - I(f) \right| \le V(f) D^*(\mathcal{P}_K), \tag{9}$$

where $I(f)$ and $\hat{I}_K(f, \mathcal{P}_K)$ are given in (7) and (8), respectively, $V(f)$ is the variation of $f(x)$ in the sense of Hardy and Krause (Fang and Wang, 1994, p. 64 [15]) and $D^*(\mathcal{P}_K)$ is the star discrepancy that measures the uniformity of distribution of a finite point set $\mathcal{P}_K = \{x_k : 1 \le k \le K\}$, see Fang and Wang (1994) for the details. The Koksma-Hlawka inequality implies that the absolute integration error is bounded by the star discrepancy of $\mathcal{P}_K$ multiplied by the variation $V(f)$. Since the latter is fixed as along as the function f is given, the absolute integration error can be minimized by carefully choosing the integration nodes $\mathcal{P}_K$. It can be shown that the absolute integration error has an order $O\left((\log K)^{q-1} / K\right)$, so that integration nodes $\mathcal{P}_K$ should be chosen such that their start discrepancy is close to $O\left((\log K)^{q-1} K^{-1}\right)$. For large K, this rate of convergence is considerably faster than the crude MC approach whose error is bounded to $O(K^{-1/2})$. More details regarding to the superiority of the QMC method over the MC method can be found in Morokoff and Caflisch (1995) [16], and Fang and Wang (1994) [15].

In the QMC approach, there are many good algorithms for generating such integration nodes. One set of such integration nodes is called good point set and discussed below.

3.3. Good Point (GP) Set

Denote a good point by $v = (v_1, v_2, \cdots, v_q) \in C^q$ where $C^q = \{(x_1, x_2, \cdots, x_q) : 0 \le x_i \le 1, i = 1, 2, \cdots, q\}$ is a hypercube. The set $\mathcal{P}_K$ consists of the first K points of the set $\left\{ (\{v_1 k\}, \{v_2 k\}, \cdots, \{v_q k\}), k = 1, 2, \cdots, K \right\}$, where $\{v_i k\}$ represents the fractional part of $v_i k$ for $i = 1, 2, \cdots, q$. In practice, the following form of the good point v is quite common to use,

$$v = \left(\left\{ \sqrt{p_1} \right\}, \left\{ \sqrt{p_2} \right\}, \cdots, \left\{ \sqrt{p_q} \right\} \right), \tag{10}$$

where p_i is a prime number and $p_i \ne p_j (i \ne j)$, $1 \le i \le q$. For example, they can be chosen as the first q primes.

3.4. MLE via QMC Approximation

Suppose $\{c_k : k = 1, 2, \cdots, K\}$ is a GP set. When the QMC method is applied to the marginal quasi-likelihood function (6), we obtain

$$\ell(\beta, \theta) = \log\left[\frac{1}{K}\sum_{k=1}^{K}\exp\left\{\sum_{i=1}^{n}\ell_i\left(\beta, \theta \Big| \Sigma_b^{\frac{1}{2}}(\theta)F^{-1}(c_k)\right)\right\}\right], \tag{11}$$

where $F^{-1}(\cdot)$ is the inverse of $F(\cdot)$, the CDF of random effects 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 $b_k = \Sigma_b^{\frac{1}{2}}F^{-1}(c_k)$, so that $\eta_{ik}^{b_k} = x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}b_k = x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}\Sigma_b^{\frac{1}{2}}F^{-1}(c_k)$ and $\mu_{ik}^{b_k} = h(\eta_{ik}^{b_k})$ where $h(\cdot)$ is the inverse function of the link function $g(\cdot)$. The maxima of $\ell(\beta, \theta)$ with respect to β and θ is the solution of the score equations

$$\frac{\partial\ell(\beta, \theta)}{\partial\beta} = \sum_{k=1}^{K}w_k\left(\sum_{i=1}^{n}\frac{a_i(y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k})\dot{g}(\mu_{ik}^{b_k})}\cdot x_i\right) = 0, \tag{12}$$

$$\frac{\partial\ell(\beta, \theta)}{\partial\theta} = \sum_{k=1}^{K}w_k\left(\sum_{i=1}^{n}\frac{a_i(y_i - \mu_{ik}^{b_k})}{\phi v(\mu_{ik}^{b_k})\dot{g}(\mu_{ik}^{b_k})}\cdot z_i'\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}(c_k)\right) = 0 \tag{13}$$

where $\dot{\Sigma}_{\theta_j}^{\frac{1}{2}} = \frac{\partial\Sigma_b^{\frac{1}{2}}}{\partial\theta_j}$, and w_k is a weight for the *k*th point given by

$$w_k = \frac{\exp\left\{\sum_{i=1}^{n}\ell_i(\beta, \theta | b_k)\right\}}{\sum_{k=1}^{K}\exp\left\{\sum_{i=1}^{n}\ell_i(\beta, \theta | b_k)\right\}}. \tag{14}$$

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

In general, the score equations in (12)-(13) have no analytical solutions for β and θ. We then propose to use Newton-Raphson algorithm to update the solutions, though calculation of Hessian matrix is somewhat difficult. In Appendices A and B of this paper, we report the detailed calculation of Hessian matrix. When the difference between $\beta^{(m)}$ and $\beta^{(m-1)}$, and/or the difference between $\theta^{(m)}$ and $\theta^{(m-1)}$, is small enough, the maximum likelihood estimates of β and θ are achieved. At convergence, the asymptotic variance-covariance matrix of the MLE $(\hat{\beta}, \hat{\theta})$ is also obtained straightforwardly through the inverse of Hessian matrix.

4. Modeling of Covariance Structures of Random Effects

Misspecification of covariance structures of random effects may cause problems of inefficient estimates of parameters in GLMMs. In some circumstances, it may lead to biased estimates of the parameters. Hence, in this paper we propose to model the covariance structure of random effects in GLMMs, rather than specify a structure to the covariance matrix Σ_b of random effects b. In a spirit of Pourahmadi (1999; 2000) [17] [18], we consider a modified Cholesky decomposition of Σ_b^{-1}, instead of Σ_b, for obtaining a statistically meaningful parametrization of the covariance matrix. The modified Cholesky decomposition removes three difficulties arising in modeling covariance structures: a) covariance matrix is usually positive definite; b) covariance matrix may be of large size; and c) the parametrization has to take into account of computational efficiency and statistical interpretation. In the past decade, an increasing attention to the decomposition was paid by many statisticians including Daniels and Pourahmadi (2002) [19], Daniels and Zhao (2003) [20], Smith and Kohn (2002) [21], Pan and MacKenzie (2003) [22], and Ye and Pan (2006) [11] among others. The key of the modified Cholesky decomposition is that a symmetric matrix Σ_b is positive definite if and only if there exist a unique unit lower tri-

angular matrix T with 1's as its diagonal entries, and a unique diagonal matrix D with positive diagonal entries, such that

$$T\Sigma_b T' = D ,\qquad(15)$$

where

$$T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{21} & 1 & 0 & \cdots & 0 \\ -\phi_{31} & -\phi_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{q1} & -\phi_{q2} & -\phi_{q3} & \cdots & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_q^2 \end{bmatrix}.\qquad(16)$$

Thus, $\Sigma_b = T^{-1}D\left(T'\right)^{-1}$ and $\Sigma_b^{-1} = T'D^{-1}T$. The modified Cholesky decomposition of Σ_b^{-1} offers a simple unconstrained and statistically meaningful reparametrization of the covariance matrix. T and D are easy to compute and interpret statistically: the below-diagonal entries of T are the negative of autoregressive coefficients (ACs), say ϕ_{jk}, in the autoregressive model

$$\hat{b}_j = \sum_{k=1}^{j-1}\phi_{jk}b_k \qquad \left(1 \le k < j \le q\right),\qquad(17)$$

where b_k is the kth component of the vector of random effects b. In other words, $\hat{b}_j$ is the linear least squares predictor of b_j based on its predecessors $b_{j-1},\cdots,b_1$. On the other hand, the diagonal entries of D are the innovation variances (IVs), *i.e.*, $\sigma_j^2 = Var\left(e_j\right)$ where $e_j = b_j - \hat{b}_j$. Obviously, $Cov\left(e_j, e_k\right) = 0$ if $j \ne k$ $\left(1 \le j, k \le q\right)$. Let $e = \left(e_1, \cdots, e_q\right)'$, then $e = Tb$ so that $e \sim N_q\left(0, D\right)$.

In a spirit of Pourahmadi (1999) [17], we propose to model the ACs and the logarithm of the IVs using linear regression models

$$\begin{cases} \phi_{jk} = c'_{jk}\gamma \\ \log\sigma_j^2 = h'_j\lambda \end{cases}\qquad(18)$$

where $c'_{jk}\left(1 \times q_1^*\right)$ and $h'_j\left(1 \times q_2^*\right)$ are covariates, and γ and λ are low-dimensional parameter vectors. Hence, the parameter vector of variance components is $\theta = \left(\gamma', \lambda'\right)'$.

5. Salamander Data Analysis

Assume that y_{ij} is the outcome for the mating of the ith female salamander with the jth male. Consider the following logit mixed model for each experiment

$$\text{logit}\left(p_{ij}^b\right) = \log\left(\frac{p_{ij}^b}{1 - p_{ij}^b}\right) = x'_{ij}\beta + b_i^f + b_j^m \qquad \left(i, j = 1, \cdots, 20\right)\qquad(19)$$

where $p_{ij}^b = Pr\left\{y_{ij} = 1 \middle| b_i^f, b_j^m\right\}$ is the conditional probability of successful mating of the pair, $x_{ij} = \left(1, ws_i^f, ws_j^m, ws_i^f \times ws_j^m\right)'$ with $ws_i^f = 1$ if the ith female comes from WS, and 0 otherwise. Similarly, $ws_j^m = 1$ if the jth male comes from WS, and 0 otherwise.

Note that each experiment involves 40 salamanders so that the log-likelihood function for each experiment involves one 40-dimensional integral. It can be further reduced to the sum of two 20-dimensional integrals due to the block design of the two closed groups, see Karim and Zeger (1992) [5] and Shun (1997) [23] for more details about the design. When the three-experiment data sets are pooled as those in model (19), the log-likelihood function $\ell\left(\beta, \theta\right)$ is a sum of six 20-dimensional integrals which are analytically intractable, so that the MLEs of the fixed effects β and variance components θ become extremely difficult to calculate. In the literature, several analytical and numerical approximations were proposed to overcome the difficulties in evaluating the log-likelihood function. Examples include MCMC method by Karim and Zeger (1992) [5], MCEM algorithm by McCulloch (1994) [6] and PQL estimation by Breslow and Clayton (1993) [2]. However, all the literature work assumed that the covariance matrix Σ_b of random effects has certain structures.

We now apply the proposed covariance modeling method to the salamander mating data. Note $q = 20$. We consider the following model for the autoregressive coefficients,

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_{2,1} \\ \phi_{3,1} \\ \phi_{3,2} \\ \vdots \\ \phi_{q,1} \\ \phi_{q,2} \\ \vdots \\ \phi_{q,q-1} \end{pmatrix} = \begin{pmatrix} c'_{2,1} \\ c'_{3,1} \\ c'_{3,2} \\ \vdots \\ c'_{q,1} \\ c'_{q,2} \\ \vdots \\ c'_{q,q-1} \end{pmatrix} \gamma = \boldsymbol{C}\gamma \tag{20}$$

where ϕ_{d_1,d_2} is the negative of the (d_1, d_2)th element of the lower triangular matrix T in the modified Cholesky decomposition ($d_1 = 2, \cdots, q$; $d_2 = 1, \cdots, q-1$), so that $\boldsymbol{\Phi}$ is a $(q-1)q/2$-dimensional vector. $\boldsymbol{C}$ is an $(q-1)q/2 \times q_1^*$ $\left(q_1^* = 4\right)$ design matrix with rows

$$c'_{d_1,d_2} = \left(1, G_{d_1,d_2}, D_{d_1,d_2}, G_{d_1,d_2} \times D_{d_1,d_2}\right) \tag{21}$$

where $G_{d_1,d_2} = 1$ if the d_1^{th} and d_2^{th} salamanders come from different genders, and 0 otherwise. Similarly, $D_{d_1,d_2} = 1$ if the d_1^{th} and d_2^{th} salamanders come from different districts, and 0 otherwise. γ is an q_1^*-variant vector of parameters.

The model for the innovation variances is

$$\boldsymbol{\Omega} = \begin{pmatrix} \log\sigma_1^2 \\ \vdots \\ \log\sigma_q^2 \end{pmatrix} = \begin{pmatrix} h'_1 \\ \vdots \\ h'_q \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{q_2^*} \end{pmatrix} = \boldsymbol{H}\lambda \tag{22}$$

where $\boldsymbol{\Omega}$ is an q-variant vector. $\boldsymbol{H}$ is an $q \times q_2^*$ design matrix $\left(q_2^* = 4\right)$ with rows

$$h'_j = \left(1, Gender_j, District_j, Gender_j \times District_j\right) \quad j = 1, \cdots, q \tag{23}$$

where $Gender_j = 1$ if the jth salamander is female, and 0 otherwise. Similarly, $District_j = 1$ if the jth salamander comes from WS, and 0 otherwise. Note λ is an q_2^*-variant vector of parameters and σ_j^2 $\left(j = 1, \cdots, q\right)$ are the diagonal elements of the matrix D in the modified Cholesky decomposition.

The pooled data are now analyzed using the proposed QMC approach. The MLEs of the fixed effects and variance components in the model are calculated. Specifically, we implement the simplest QMC points, square root sequences, to approximate the 20-dimensional analytically intractable integrals. A set of 20-dimensional points on the unit cube $C^{20} = [0,1)^{20}$ are generated to approximate the six 20-dimensional integrals in the marginalized log-likelihood function $\ell(\beta,\theta)$. We then use Newton-Raphson algorithm to maximize the approximated log-likelihood function for the pooled data.

Note that the covariance modeling for $\Sigma_b = Cov(b)$, the covariance matrix of random effects, uses $q_1^* = 4$ parameters for modeling of the autoregressive coefficients and $q_2^* = 4$ parameters for the innovation variances. By using the joint mean-covariance modeling strategy we can model the structure of Σ_b and estimate the covariance matrix without specifying a covariance structure to Σ_b. This avoids the problem of misspecification of covariance structures. In addition, the number of parameters in Σ_b reduces from $q(q+1)/2$ to $q_1^* + q_2^*$, which avoids high-dimensional problems of covariance estimation when q is large. Finally, the modeling strategy based on the modified Cholesky decomposition guarantees that the resulting estimate of Σ_b is positive definite.

We code our programme in FORTRAN and run on a PC Pentium (R) 4 PC (CPU 3.20 GHz). We choose the PQL estimate as the starting value of β, i.e., $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)' = (0.79, -2.29, -0.54, 2.82)'$, and zero as for $\theta = (\gamma', \lambda')'$. **Table 1** reports the numerical results when increasing the number of integration nodes generated by the square root sequences. From **Table 1**, we can see that the simplest GP point set works reasonably well for calculating the MLEs of the parameters in models (19), (20) and (22). The results show that the maximized

Table 1. MLEs of the parameters via covariance modelling for the pooled salamander mating data using square root sequences by varying *K*, the number of square root sequence points (standard errors in parentheses).

K	β_0	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	λ_1	λ_2	λ_3	λ_4	$\hat{l}_{max}$
10,000	0.91	−2.94	−0.62	3.71	0.00045	0.00062	0.0044	0.0031	0.19	−0.022	−0.024	0.00021	−207.21
	(0.38)	(0.39)	(0.33)	(0.12)	(0.0024)	(0.0021)	(0.0056)	(0.0067)	(0.046)	(0.051)	(0.025)	(0.0046)	
20,000	0.77	−2.87	−0.59	3.58	0.00087	0.0019	0.0052	0.0039	0.064	0.0081	−0.0024	−0.00010	−207.43
	(0.34)	(0.52)	(0.50)	(0.52)	(0.0056)	(0.0055)	(0.0089)	(0.0087)	(0.034)	(0.044)	(0.042)	(0.0018)	
30,000	1.24	−2.88	−1.00	3.71	0.00030	0.0035	−0.00082	−0.0063	0.43	0.0050	0.00047	0.00054	−205.95
	(0.15)	(0.37)	(0.22)	(0.57)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0019)	(0.026)	(0.023)	(0.0027)	
40,000	1.20	−2.84	−0.96	3.64	−0.00032	0.00092	−0.00091	0.0027	0.39	0.0031	0.0032	0.00012	−206.34
	(0.32)	(0.48)	(0.25)	(0.58)	(0.00026)	(0.0000)	(0.0000)	(0.0000)	(0.031)	(0.025)	(0.031)	(0.0033)	
50,000	1.20	−2.82	−1.00	3.70	0.00021	0.00093	0.000042	0.0014	0.43	0.00012	0.0029	0.00095	−206.46
	(0.31)	(0.46)	(0.27)	(0.54)	(0.0000)	(0.00024)	(0.0000)	(0.0000)	(0.0022)	(0.024)	(0.019)	(0.0000)	
60,000	1.16	−2.84	−0.97	3.68	0.00023	0.00089	−0.000053	0.00067	0.41	−0.00029	0.0017	−0.00036	−206.63
	(0.33)	(0.37)	(0.30)	(0.53)	(0.00017)	(0.0000)	(0.00045)	(0.00045)	(0.016)	(0.040)	(0.032)	(0.00064)	
70,000	1.22	−2.85	−0.95	3.72	0.00011	0.00057	−0.00021	−0.0023	0.52	−0.00034	0.0010	−0.00087	−206.75
	(0.30)	(0.49)	(0.18)	(0.57)	(0.00022)	(0.0000)	(0.00023)	(0.00054)	(0.035)	(0.055)	(0.025)	(0.00026)	
80,000	1.21	−2.91	−0.93	3.73	−0.00091	−0.0034	0.0045	−0.0032	0.51	0.00091	0.00024	0.00020	−206.25
	(0.40)	(0.41)	(0.15)	(0.49)	(0.00087)	(0.00095)	(0.00066)	(0.0037)	(0.023)	(0.026)	(0.014)	(0.00034)	
90,000	1.16	−2.86	−0.87	3.71	−0.00034	−0.0035	0.0054	−0.0086	0.50	−0.0027	−0.0023	0.00041	−206.94
	(0.23)	(0.45)	(0.22)	(0.60)	(0.0013)	(0.0037)	(0.0034)	(0.0000)	(0.00031)	(0.029)	(0.026)	(0.00091)	
100,000	1.25	−2.88	−0.91	3.68	−0.00042	−0.0022	0.0031	−0.0032	0.45	−0.0022	−0.00020	0.00034	−206.98
	(0.41)	(0.52)	(0.17)	(0.63)	(0.0014)	(0.00058)	(0.00010)	(0.0063)	(0.018)	(0.025)	(0.020)	(0.00062)	

log-likelihood $\hat{\ell}_{max}$ changes very little when the number of integration nodes increases from 10,000 to 100,000, implying that any number of the points between these numbers would be fine to obtain reasonably good estimates of the parameters. On the other hand, the parameter estimates using these sequences become stable quickly. We notice that the integral space is 20-dimensional so that the points are very sparse even if the number of the QMC points increases to $K = 100,000$. In the meantime, the algorithm gets convergence very quickly and it only takes a few of iterations to achieve the convergence. In the case of using $K = 100,000$ integration nodes, for example, our Fortran code takes only several minutes to obtain the estimation results.

It is noted that in the autoregressive coefficients model, the estimates of γ, including $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{\gamma}_3$ and $\hat{\gamma}_4$, are all very close to zero and not statistically significant. Therefore, the estimates of the autoregressive coefficients can be regarded as zero. On the other hand, in the model of the log-innovation variances, the estimates of the regression parameters λ, except $\hat{\lambda}_1$, are not statistically significant.

In other words, *T* is an identity matrix and *D* is a diagonal matrix with an identical element on diagonals, implying $\hat{\Sigma}_b = \hat{\sigma}_b^2 I_q$ where $\hat{\sigma}_b^2 = \exp\{\hat{\lambda}_1\}$. Hence, it is concluded that the random effects in the modeling of the salamander mating data are uncorrelated, and so there is no heterogeneity between the female and male salamanders.

By comparing the numerical results of the QMC estimation approach with the literature work, we find that they are quite close to those made by Gibbs sampling method. But the QMC method has much light computational loads and easy to use for practitioners. This is because Gibbs sampling method involves multiple draws of random samples and needs more experience, for instance, in specifying prior distributions of parameters, etc., whereas the QMC approximation method does not need such specifications. Also, Gibbs sampling may be very

slow in obtaining the parameter estimates. On the other hand, the PQL approach gives very biased estimates of the variance components for modeling clustered binary data, as reported by many authors including Breslow and Clayton (1993) [2], Breslow and Lin (1995) [24], Booth and Hobert (1999) [25] and Aiken (1999) [26].

6. Simulation

In this section, we carry out simulation studies to assess the performance of the proposed QMC estimation method in GLMMs. In the simulation studies we consider the logistic model (19) and use the same protocol of design as the real salamander mating experiment, resulting in 360 correlated binary observations. We run 100 simulations and calculate the average of the parameter estimates over the 100 simulations. The log-likelihood function for each simulated data set involves six 20-dimensional integrals that are analytically intractable.

We generate $K = 50,000$ integration nodes on the unit cube $C^{20} = [0,1)^{20}$ using the square root sequences. We then use those integration nodes to approximate the integrated log-likelihood as the one given in (11), and use Newton-Raphson algorithm to maximize the approximated log-likelihood function. The true fixed effects β are chosen to be $\beta = (1.2, -2.8, -1, 3.6)'$, which is quite close to the one obtained in the real data analysis presented in the previous section. In the real data analysis there is no heterogeneity between female and male, but in our simulation studies we choose $\gamma = (-1.2, 0.6, -0.1, 0.4)'$ for modeling the ACs and $\lambda = (-2, 1, 0.5)'$ for the IVs in order to assess the performance of the proposed method when heterogeneity occurs. The starting values of the parameters are chosen to be the parameter estimates from the real data analysis. The estimation results of the simulation studies are summarized in **Table 2**, which presents the average of the parameter estimates over 100 replications and their standard errors as well.

From **Table 2** it is clear that the proposed QMC approach is able to produce reasonably good estimates of parameters in GLMMs even if there are heterogenous random effects. The estimates of the fixed effects and variance components, based on the average of 100 simulation results, are quite close to the true values of the parameters. It implies that the proposed QMC approach with covariance modeling strategy works well in estimating parameters in GLMMs with correlated random effects.

7. Discussions

We have studied the performance of using QMC approach to calculate the MLEs of the fixed effects and variance components in GLMMs with correlated random effects. We proposed to use Newton-Raphson algorithm to calculate the parameter estimates. The marginalized log-likelihood function that is in general analytically intractable is approximated well through the use of the simplest QMC points, *i.e.*, square root sequences. We also addressed the issue of covariance modelling for random effects covariance matrix through a modified Cholesky decomposition. As a result, pre-specification of covariance structures for random effects is not necessary and misspecification of covariance structures is thus avoided. The score function and observed information matrix are calculated and expressed in analytically closed forms, so that the algorithm can be implemented straightforwardly. The performance of the proposed method was assessed through real salamander mating data analysis and simulation studies. Even though the marginalized log-likelihood function in the numerical analysis involves six 20-dimensional analytically intractable integrals, the QMC square root sequence approximation performs very well.

Table 2. Average of 100 parameter estimates in the simulation studies, where the QMC approximation uses $K = 50,000$ points implemented using the square root sequences (simulated standard errors are provided 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
QMC-gp	1.16	−2.82	−0.93	3.65	−1.21	0.55	−0.12	0.41
StD	0.04	0.12	0.08	0.17	0.000094	0.0027	0.0020	0.016
	λ_1	λ_2	λ_3					
True	−2.00	−1.00	0.50					
QMC-gp	−1.99	−1.03	0.51					
StD	0.0028	0.0017	0.0017					

In general, a practical issue for the use of the QMC points is the choice of the number of integration nodes. Pan and Thompson (2007) [10] suggested to increase the number of integration points gradually until the estimation results become relatively stable. This is exactly what we have done in the numerical analysis displayed in **Table 1**.

Acknowledgements

References

[1] 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

[2] Breslow, N.E. and Clayton, D.G. (1993) Approximate Inference Ingeneralized Linear Mixed Models. *Journal of the American Statistical Association*, **88**, 9-25.

[3] Lin, X. and Breslow, N.E. (1996) Bias Correction in Generalized Lnear 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

[4] Lee, Y. and Nelder, J.A. (2001) Hierrarchical 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

[5] 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

[6] McCulloch, C.E. (1994) Maximum Likelihood Variance Components Estimation for Binary Data. *Journal of the American Statistical Association*, **89**, 330-335. http://dx.doi.org/10.1080/01621459.1994.10476474

[7] McCulloch, C.E. (1997) Maximum Likelihood Algorithms for Generallized Linear Mixed Models. *Journal of the American Statistical Association*, **92**, 162-170. http://dx.doi.org/10.1080/01621459.1997.10473613

[8] Chan, J.S., Kuk, A.Y. and Yam, C.H. (2005) Monte Carlo Approximation through Gibbs Output in Generalized Linear Mixed Models. *Journal of Maltivariate Analysis*, **94**, 300-312. http://dx.doi.org/10.1016/j.jmva.2004.05.004

[9] Pan, J. and Thompson, R. (2003) Gauss-Hermite Quadrature Approximation for Estimation in Generalised Linear Mixed Models. *Computational Statistics*, **18**, 57-78.

[10] 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

[11] 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

[12] Niederreiter, H. (1992) Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia. http://dx.doi.org/10.1137/1.9781611970081

[13] Spanier, J. and Maize, E. (1994) Quasi-Random Methods for Estimating Integrals Using Relatively Small Samples. *SIAM Review*, **36**, 19-44. http://dx.doi.org/10.1137/1036002

[14] Traub, J.F. and Wozniakowski, H. (1992) The Monte Carlo Algorithm with Pseudorandom Generator. *Mathematics of Computation*, **58**, 323-339. http://dx.doi.org/10.1090/S0025-5718-1992-1106984-4

[15] 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

[16] Morokoff, W.J. and Caflisch, R.E. (1995) Quasi-Monte Carlo Integration. *Journal of Computational Physics*, **122**, 218-230. http://dx.doi.org/10.1006/jcph.1995.1209

[17] 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

[18] Pourahmadi, M. (2000) Maximum Likelihood Estimation of Generalised Linear Models for Multivariate Normal Covariance Matrix. *Biometrika*, **87**, 425-435. http://dx.doi.org/10.1093/biomet/87.2.425

[19] Daniels, M.J. and Pourahmadi, M. (2002) Bayesian Analysis of Covariance Matrices and Dynamic Models for Longitudinal Data. *Biometrika*, **89**, 553-566. http://dx.doi.org/10.1093/biomet/89.3.553

[20] Daniels, M.J. and Zhao, Y.D. (2003) Modelling the Random Effects Covariance Matrix in Longitudinal Data. *Statistics*

in Medicine, **22**, 1631-1647. http://dx.doi.org/10.1002/sim.1470

[21] Smith, M. and Kohn, R. (2002) Parsimonious Covariance Matrix Estimation for Longitudinal Data. *Journal of the American Statistical Association*, **97**, 1141-1153. http://dx.doi.org/10.1198/016214502388618942

[22] Pan, J. and MacKenzie, G. (2003) On Modelling Mean-Covariance Structures in Longitudinal Studies. *Biometrika*, **90**, 239-244. http://dx.doi.org/10.1093/biomet/90.1.239

[23] 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

[24] Breslow, N.E. and Lin, X. (1995) Bias Correction in Generalised Linear Mixed Models with a Single-Component of Dispersion. *Biometrika*, **82**, 81-91. http://dx.doi.org/10.1093/biomet/82.1.81

[25] 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*, *B*, **61**, 265-285. http://dx.doi.org/10.1111/1467-9868.00176

[26] Aitken, M. (1999) A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models. *Biometrics*, **55**, 117-128. http://dx.doi.org/10.1111/j.0006-341X.1999.00117.x

Appendix A: Second-Order Derivatives of Log-Likelihood

The second-order derivatives of log-likelihood function is given by

$$\frac{\partial^2 l_i}{\partial \beta \partial \beta^{\mathrm{T}}} = \frac{\partial}{\partial \beta}\left(\frac{\partial l_i}{\partial \beta^{\mathrm{T}}}\right) = \frac{\partial}{\partial \beta}\left[\left(\frac{\partial l_i}{\partial \beta}\right)^{\mathrm{T}}\right] = \frac{\partial}{\partial \beta}\left[\frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}x_i^{\mathrm{T}}\right] = \frac{a_i}{\phi}\frac{\partial}{\partial \beta}\left(\frac{y_i - \mu_{ik}^{b_k}}{v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}\right)x_i^{\mathrm{T}}$$

$$= \frac{a_i}{\phi}\frac{1}{v^2\dot{g}^3}\left[-v\dot{g} - \left(y_i - \mu_{ik}^{b_k}\right)\left(\dot{v}\dot{g} + \ddot{g}v\right)\right]x_i x_i^{\mathrm{T}}$$

$$\frac{\partial^2 l_i}{\partial \theta_j \partial \theta_j'} = \frac{\partial}{\partial \theta_j}\left(\frac{\partial l_i}{\partial \theta_j'}\right) = \frac{\partial}{\partial \theta_j}\left[\left(\frac{\partial l_i}{\partial \theta_j}\right)^{\mathrm{T}}\right] = \frac{\partial}{\partial \theta_j}\left(\frac{\partial l_i}{\partial \theta_j}\right) = \frac{\partial^2 l_i}{\partial \theta_j^2} = \frac{\partial}{\partial \theta_j}\left[\frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}z_i'\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}\left(c_k\right)\right]$$

$$= \frac{a_i}{\phi}\frac{\partial}{\partial \theta_j}\left(\frac{y_i - \mu_{ik}^{b_k}}{v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}\right)z_i'\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}\left(c_k\right) + \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}z_i'\ddot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}\left(c_k\right)$$

$$= \frac{a_i}{\phi}\frac{1}{v^2\dot{g}^3}\left[-v\dot{g} - \left(y_i - \mu_{ik}^{b_k}\right)\left(\dot{v}\dot{g} + \ddot{g}v\right)\right]\left(z_i'\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}\left(c_k\right)\right)^2 + \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}z_i'\ddot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}\left(c_k\right)$$

$$\frac{\partial^2 l_i}{\partial \theta_j \partial \theta_k'} = \frac{\partial}{\partial \theta_j}\left(\frac{\partial l_i}{\partial \theta_k'}\right) = \frac{\partial}{\partial \theta_j}\left(\frac{\partial l_i}{\partial \theta_k}\right) = \frac{\partial^2 l_i}{\partial \theta_j \partial \theta_k} = \frac{\partial}{\partial \theta_j}\left[\frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}z_i'\dot{\Sigma}_{\theta_k}^{\frac{1}{2}}F^{-1}\left(c_k\right)\right]$$

$$= \frac{\partial}{\partial \theta_j}\left(\frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}\right)z_i'\dot{\Sigma}_{\theta_k}^{\frac{1}{2}}F^{-1}\left(c_k\right) + \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}\cdot\frac{\partial}{\partial \theta_j}\left(z_i'\dot{\Sigma}_{\theta_k}^{\frac{1}{2}}F^{-1}\left(c_k\right)\right)$$

$$= \frac{a_i}{\phi}\frac{1}{v^2\dot{g}^3}\left[-v\dot{g} - \left(y_i - \mu_{ik}^{b_k}\right)\left(\dot{v}\dot{g} + \ddot{g}v\right)\right]\left(z_i'\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}\left(c_k\right)\right)\left(z_i'\dot{\Sigma}_{\theta_k}^{\frac{1}{2}}F^{-1}\left(c_k\right)\right)$$

$$+ \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}z_i'\ddot{\Sigma}_{\theta_j\theta_k}^{\frac{1}{2}}F^{-1}\left(c_k\right)$$

$$\frac{\partial^2 l_i}{\partial \beta \partial \theta_j} = \frac{\partial}{\partial \beta}\left(\frac{\partial l_i}{\partial \theta_j}\right) = \frac{\partial}{\partial \beta}\left[\frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)}z_i'\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}\left(c_k\right)\right]$$

$$= \frac{a_i}{\phi}\frac{1}{v^2\dot{g}^3}\left[-v\dot{g} - \left(y_i - \mu_{ik}^{b_k}\right)\left(\dot{v}\dot{g} + \ddot{g}v\right)\right]x_i z_i'\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}\left(c_k\right)$$

$$\frac{\partial w_k}{\partial \beta} = w_k\left(\sum_{i=1}^{n}\frac{\partial \ell_i\left(\beta,\theta|b_k\right)}{\partial \beta}\right) - w_k\sum_{k=1}^{K}w_k\left(\sum_{i=1}^{n}\frac{\partial \ell_i\left(\beta,\theta|b_k\right)}{\partial \beta}\right) = w_k\left(\sum_{i=1}^{n}\frac{\partial \ell_i}{\partial \beta}\right) - w_k\frac{\partial \ell}{\partial \beta}$$

$$\frac{\partial w_k}{\partial \theta_j} = w_k\left(\sum_{i=1}^{n}\frac{\partial \ell_i\left(\beta,\theta|b_k\right)}{\partial \theta_j}\right) - w_k\sum_{k=1}^{K}w_k\left(\sum_{i=1}^{n}\frac{\partial \ell_i\left(\beta,\theta|b_k\right)}{\partial \theta_j}\right) = w_k\left(\sum_{i=1}^{n}\frac{\partial \ell_i}{\partial \theta_j}\right) - w_k\frac{\partial \ell}{\partial \theta_j}$$

$$\frac{\partial^2 l}{\partial \beta \partial \beta^{\mathrm{T}}} = \frac{\partial}{\partial \beta}\left(\frac{\partial l}{\partial \beta^{\mathrm{T}}}\right) = \frac{\partial}{\partial \beta}\left[\left(\frac{\partial l}{\partial \beta}\right)^{\mathrm{T}}\right] = \frac{\partial}{\partial \beta}\left[\sum_{k=1}^{K} w_k \left(\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} x_i^{\mathrm{T}}\right)\right]$$

$$= \sum_{k=1}^{K} w_k \left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} x_i\right]\left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} x_i^{\mathrm{T}}\right]$$

$$- \left[\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} x_i\right]\left[\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} x_i^{\mathrm{T}}\right]$$

$$- \sum_{k=1}^{K} w_k \left(\sum_{i=1}^{n} \frac{a_i}{\phi}\left\{\left[v\left(\mu_{ik}^{b_k}\right)\dot{g}^2\left(\mu_{ik}^{b_k}\right)\right]^{-1} + \left[\frac{\dot{v}\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right) + v\left(\mu_{ik}^{b_k}\right)\ddot{g}\left(\mu_{ik}^{b_k}\right)}{v^2\left(\mu_{ik}^{b_k}\right)\dot{g}^3\left(\mu_{ik}^{b_k}\right)}\right]\left(y_i - \mu_{ik}^{b_k}\right)\right\} x_i x_i^{\mathrm{T}}\right)$$

$$\frac{\partial^2 l}{\partial \theta_j \partial \theta_j} = \frac{\partial}{\partial \theta_j}\left(\frac{\partial l}{\partial \theta_j}\right) = \frac{\partial}{\partial \theta_j}\left[\sum_{k=1}^{K} w_k \left(\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right)\right]$$

$$= \sum_{k=1}^{K} w_k \left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right]\left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right]$$

$$- \left[\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right]^2$$

$$- \sum_{k=1}^{K} w_k \left(\sum_{i=1}^{n} \frac{a_i}{\phi}\left\{\left[v\left(\mu_{ik}^{b_k}\right)\dot{g}^2\left(\mu_{ik}^{b_k}\right)\right]^{-1} + \left[\frac{\dot{v}\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right) + v\left(\mu_{ik}^{b_k}\right)\ddot{g}\left(\mu_{ik}^{b_k}\right)}{v^2\left(\mu_{ik}^{b_k}\right)\dot{g}^3\left(\mu_{ik}^{b_k}\right)}\right]\right\}\left(z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right)^2\right)$$

$$+ \sum_{k=1}^{K} w_k \left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \ddot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right]$$

$$\frac{\partial^2 l}{\partial \theta_j \partial \theta_k} = \frac{\partial}{\partial \theta_j}\left(\frac{\partial l}{\partial \theta_k}\right) = \frac{\partial}{\partial \theta_j}\left[\sum_{k=1}^{K} w_k \left(\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_k}^{\frac{1}{2}} F^{-1}(c_k)\right)\right]$$

$$= \sum_{k=1}^{K} w_k \left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right]\left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_k}^{\frac{1}{2}} F^{-1}(c_k)\right]$$

$$- \left[\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right]\left[\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_k}^{\frac{1}{2}} F^{-1}(c_k)\right]$$

$$- \sum_{k=1}^{K} w_k \left(\sum_{i=1}^{n} \frac{a_i}{\phi}\left\{\left[v\left(\mu_{ik}^{b_k}\right)\dot{g}^2\left(\mu_{ik}^{b_k}\right)\right]^{-1} + \left[\frac{\dot{v}\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right) + v\left(\mu_{ik}^{b_k}\right)\ddot{g}\left(\mu_{ik}^{b_k}\right)}{v^2\left(\mu_{ik}^{b_k}\right)\dot{g}^3\left(\mu_{ik}^{b_k}\right)}\right]\right\}\left(z_i^{\mathrm{T}} \dot{\Sigma}_{\theta_j}^{\frac{1}{2}} F^{-1}(c_k)\right)^2\right)$$

$$+ \sum_{k=1}^{K} w_k \left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}} \ddot{\Sigma}_{\theta_j \theta_k}^{\frac{1}{2}} F^{-1}(c_k)\right]$$

$$\frac{\partial^2 l}{\partial \beta \partial \theta_j} = \frac{\partial}{\partial \beta}\left(\frac{\partial l}{\partial \theta_j}\right) = \frac{\partial}{\partial \beta}\left[\sum_{k=1}^{K} w_k \left(\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}}\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}(c_k)\right)\right]$$

$$= \sum_{k=1}^{K} w_k \left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}}\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}(c_k)\right]\left[\sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} x_i\right]$$

$$- \left[\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} z_i^{\mathrm{T}}\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}(c_k)\right]\left[\sum_{k=1}^{K} w_k \sum_{i=1}^{n} \frac{a_i\left(y_i - \mu_{ik}^{b_k}\right)}{\phi v\left(\mu_{ik}^{b_k}\right)\dot{g}\left(\mu_{ik}^{b_k}\right)} x_i\right]$$

$$- \sum_{k=1}^{K} w_k \left(\sum_{i=1}^{n} \frac{a_i}{\phi}\left[v\left(\mu_{ik}^{b_k}\right)\dot{g}^2\left(\mu_{ik}^{b_k}\right)\right]^{-1} z_i^{\mathrm{T}}\dot{\Sigma}_{\theta_j}^{\frac{1}{2}}F^{-1}(c_k)\cdot x_i\right)$$

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

Appendix B: MLEs for Covariance Modeling

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

$$\dot{\Sigma}_{\gamma_j}^{\frac{1}{2}} = \frac{\partial \Sigma_b^{\frac{1}{2}}}{\partial \gamma_j} = \frac{\partial T^{-1}(\gamma)}{\partial \gamma_j}\cdot D^{\frac{1}{2}}(\lambda) \quad \left(j = 1,\cdots,q_1^*\right) \tag{24}$$

$$\ddot{\Sigma}_{\gamma_j}^{\frac{1}{2}} = \frac{\partial^2 \Sigma_b^{\frac{1}{2}}}{\partial \gamma_j^2} = \frac{\partial^2 T^{-1}(\gamma)}{\partial \gamma_j^2}\cdot D^{\frac{1}{2}}(\lambda) \quad \left(j = 1,\cdots,q_1^*\right) \tag{25}$$

$$\ddot{\Sigma}_{\gamma_j\gamma_k}^{\frac{1}{2}} = \frac{\partial^2 \Sigma_b^{\frac{1}{2}}}{\partial \gamma_j \partial \gamma_k} = \frac{\partial^2 T^{-1}(\gamma)}{\partial \gamma_j \partial \gamma_k}\cdot D^{\frac{1}{2}}(\lambda) \quad \left(j \neq k \text{ and } 1 \leq j,k \leq q_1^*\right) \tag{26}$$

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

$$\Phi^* = \begin{pmatrix} 1 & & & & & \\ \phi_{21} & 1 & & & & \\ \phi_{31} & \phi_{32} & 1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ \vdots & \vdots & \vdots & & \ddots & \\ \phi_{q,1} & \phi_{q,2} & \phi_{q,3} & \cdots & \phi_{q,q-1} & 1 \end{pmatrix} \tag{27}$$

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

$$\Phi = \begin{pmatrix} \phi_{21} \\ \phi_{31} \\ \phi_{32} \\ \vdots \\ \phi_{q,q-1} \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_i \\ \vdots \\ \phi_{\frac{q(q-1)}{2}} \end{pmatrix} = \begin{pmatrix} c_{1,1} & \cdots & c_{1,j} & \cdots & c_{1,q_1^*} \\ \vdots & & \vdots & & \vdots \\ c_{i,1} & \cdots & c_{i,j} & \cdots & c_{i,q_1^*} \\ \vdots & & \vdots & & \vdots \\ c_{\frac{q(q-1)}{2},1} & \cdots & c_{\frac{q(q-1)}{2},j} & \cdots & c_{\frac{q(q-1)}{2},q_1^*} \end{pmatrix}\begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_j \\ \vdots \\ \gamma_{q_1^*} \end{pmatrix} \tag{28}$$

We can put the jth column of $\boldsymbol{C}$ into a $q \times q$ lower triangular matrix like $\boldsymbol{\Phi}^*$. So we have q_1^* $q \times q$ matrices. The element $c_{i,j}$ in (28) can be put into the place (d_1, d_2) in the jth matrix. For the jth matrix, $j = 1, \cdots, q_1^*$, we have the relationship $i = 0 + 1 + 2 + \cdots + (d_1 - 2) + d_2$ where $1 \le d_1 \le q$, $1 \le d_2 \le d_1 - 1$ and $1 \le i \le \dfrac{q(q-1)}{2}$.

The above relationship is a one-to-one correspondence between i and (d_1, d_2). So ϕ_i and $c_{i,j}$ in (28) can be also expressed as

$$\phi_i \Leftrightarrow \phi_{d_1, d_2} \qquad c_{i,j} \Leftrightarrow c_{d_1, d_2}^j \tag{29}$$

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

$$a_{d_1, d_2} = \begin{cases} \phi_{d_1 d_2} & d_1 = d_2 + 1 \\ \phi_{d_1 d_2} + \displaystyle\sum_{m = d_2 + 1}^{d_1 - 1} \phi_{d_1 m} a_{m d_2} & d_1 \ge d_2 + 2 \end{cases} \tag{30}$$

The first derivative can be calculated as

$$\frac{\partial T^{-1}}{\partial \gamma_j} = -T^{-1} \cdot \frac{\partial T}{\partial \gamma_j} \cdot T^{-1} \tag{31}$$

where $\dfrac{\partial T}{\partial \gamma_j}$ 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

$$\frac{\partial^2 T^{-1}}{\partial \gamma_j^2} = -\left[\frac{\partial T^{-1}}{\partial \gamma_j} \cdot \frac{\partial T}{\partial \gamma_j} \cdot T^{-1} + T^{-1} \cdot \frac{\partial T}{\partial \gamma_j} \cdot \frac{\partial T^{-1}}{\partial \gamma_j} \right] \tag{32}$$

$$\frac{\partial^2 T^{-1}}{\partial \gamma_j \partial \gamma_k} = -\left[\frac{\partial T^{-1}}{\partial \gamma_k} \cdot \frac{\partial T}{\partial \gamma_j} \cdot T^{-1} + T^{-1} \cdot \frac{\partial T}{\partial \gamma_j} \cdot \frac{\partial T^{-1}}{\partial \gamma_k} \right] \tag{33}$$

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

$$\dot{\Sigma}_{b \lambda_j}^{\frac{1}{2}} = \frac{\partial \Sigma_b^{\frac{1}{2}}}{\partial \lambda_j} = T^{-1}(\gamma) \cdot \frac{\partial D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j} \quad \left(j = 1, \cdots, q_2^* \right) \tag{34}$$

$$\ddot{\Sigma}_{b \lambda_j}^{\frac{1}{2}} = \frac{\partial^2 \Sigma_b^{\frac{1}{2}}}{\partial \lambda_j^2} = T^{-1}(\gamma) \cdot \frac{\partial^2 D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j^2} \quad \left(j = 1, \cdots, q_2^* \right) \tag{35}$$

$$\ddot{\Sigma}_{b \lambda_j \lambda_k}^{\frac{1}{2}} = \frac{\partial^2 \Sigma_b^{\frac{1}{2}}}{\partial \lambda_j \partial \lambda_k} = T^{-1}(\gamma) \cdot \frac{\partial^2 D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j \partial \lambda_k} \quad \left(j \ne k \text{ and } 1 \le j, k \le q_2^* \right) \tag{36}$$

where

$$D^{\frac{1}{2}}(\lambda) = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_q \end{pmatrix} = \begin{pmatrix} \exp\left(\frac{1}{2}h_1'\lambda\right) & & \\ & \ddots & \\ & & \exp\left(\frac{1}{2}h_q'\lambda\right) \end{pmatrix} \tag{37}$$

$$\frac{\partial D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j} = \frac{1}{2}\begin{pmatrix} \sigma_1 h_{1j} & & & \\ & \sigma_2 h_{2j} & & \\ & & \ddots & \\ & & & \sigma_q h_{q,j} \end{pmatrix} = \frac{1}{2}D^{\frac{1}{2}}(\lambda)\begin{pmatrix} h_{1j} & & & \\ & h_{2j} & & \\ & & \ddots & \\ & & & h_{q,j} \end{pmatrix} \tag{38}$$

$$\frac{\partial^2 D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j^2} = \frac{1}{2}\frac{\partial D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j}\begin{pmatrix} h_{1j} & & & \\ & h_{2j} & & \\ & & \ddots & \\ & & & h_{q,j} \end{pmatrix} = \frac{1}{4}D^{\frac{1}{2}}(\lambda)\begin{pmatrix} h_{1j}^2 & & & \\ & h_{2j}^2 & & \\ & & \ddots & \\ & & & h_{q,j}^2 \end{pmatrix} \tag{39}$$

$$\frac{\partial^2 D^{\frac{1}{2}}(\lambda)}{\partial \lambda_j \partial \lambda_k} = \frac{1}{2}\frac{\partial D^{\frac{1}{2}}(\lambda)}{\partial \lambda_k}\begin{pmatrix} h_{1j} & & & \\ & h_{2j} & & \\ & & \ddots & \\ & & & h_{q,j} \end{pmatrix} = \frac{1}{4}D^{\frac{1}{2}}(\lambda)\begin{pmatrix} h_{1j}h_{1k} & & & \\ & h_{2j}h_{2k} & & \\ & & \ddots & \\ & & & h_{q,j}h_{q,k} \end{pmatrix} \tag{40}$$