Modelling covariance structure in bivariate marginal models for longitudinal data

By JING XU

Centre of Biostatistics, Department of Mathematics and Statistics,

University of Limerick, Ireland

jing.xu@ul.ie

AND GILBERT MACKENZIE

ENSAI, Rennes, France and The Centre for Biostatistics,

Department of Mathematics and Statistics, University of Limerick, Ireland

gilbert.mackenzie@ul.ie

Summary

It can be more challenging and demanding to efficiently model the covariance matrices for multivariate longitudinal data than for univariate case because of the correlations between responses arising from multiple variables and repeated measurements over time. In addition to the more complicated covariance structures, the positive-definiteness constraint is still the major obstacle in modelling covariance matrices as in univariate case. In this paper, we develop a data-based method to model the covariance structures. Using this method, the constrained and hard-to-model parameters of Σ_i are traded in for unconstrained and interpretable parameters. Estimates of these parameters, together with the parameters in the mean, are obtained by maximum likelihood approach, and the large-sample asymptotic properties are derived when the observations are normally distributed. A simulation is carried out to illustrate the asymptotics. Application to a set of bivariate visual data shows that our method performs very well even when modelling bivariate nonstationary dependence structures.

Some key words: longitudinal data; bivariate marginal models; covariance modelling; block triangular factorization; matrix logarithm; modelling log-innovation matrices

1. Introduction

In many epidemiological studies and clinical trials, subjects are measured on several occasions with regard to a collection of response variables. Consider, as an example, a randomized controlled trial of Teletherapy for age-related Macular Degeneration (Hart, et. al 2002) carried out in three United-Kingdom based hospital units. 203 patients were randomly assigned to radiotherapy or observation and scheduled to visit the clinics at 0, 3, 6, 12, and 24 months. Three visual responses, distance visual acuity, near visual acuity and contrast sensitivity were taken for every patient through out the study. There are many examples of this kind (Jones, 1993; Fang et al., 2006; Gao et al., 2006).

Modelling the covariance structures for such multivariate longitudinal data are usually more complicated than for univariate case because of the occurrence of a) the correlation between the responses at each time point, b) the correlation within separate responses over time, and c) the cross-correlation between different responses at different times. Two approaches are commonly adopted in practice: models with a Kronecker product covariance structure and multivariate mixed models with random coefficients. These two approaches select the covariance structures from a limited set of potential candidate structures containing compound-symmetry, AR(1) and unstructured covariance and very often assume that the data are sampled from multiple stationary stochastic processes (Jones, 1993; Galecki, 1994; Reeves & MacKenzie, 1998; Fang et al., 2006; Gao et al., 2006).

In addition to the more complicated covariance structures, as in univariate case, the positive-definiteness is still the major obstacle in modelling the covariance matrices for multivariate longitudinal data. In the context of other types of data such as univariate longitudinal data and multivariate data, several decomposition and transformation methods are developed in the literature to partially alleviate this problem. Of these methods, frequently used are as follows: Variance-covariance decomposition (Barnard et al. 2000), Spectral decompositions (Flury, 1984; Bensmail et al., 1997; Boik, 2002, 2003), Matrix-logarithmic covariance model (Chiu et al., 1996; Pinheiro & Bates, 1996) and Cholesky Decomposition (Pourahmadi 1999; Smith & Kohn, 2002; Chen & Dunson,2003)

In this paper, we develop a data-based method to model covariance structures for bivariate longitudinal data by extending the ideas of modified Cholesky decomposition (Pourahmadi, 1999) and matrix-logarithmic covariance model (Chiu et al., 1996). There are three main steps in our method. Firstly, by block triangular factorization, the positive definite matrix $\Sigma_i = Cov(y_i)(i = 1, \cdots, n)$ is decomposed into generalized autoregressive coefficient matrices and innovation covariance matrices. These components have simpler structures and useful statistical interpretations (see subsection 2.2). Secondly, matrix-logarithmic transformation is applied to the innovation covariance matrices to ensure the resulting estimate positive definite. These new matrices, named as log-innovation covariance matrices, maintain some of the statistical features (see subsection 2.3) associated with the innovation covariance matrices. Thirdly, we model the new parameters in these matrices parsimoniously in terms of covariates using regression models (see subsection 2.4).

Thus, the constrained and hard-to-model parameters of Σ_i are traded in for the unconstrained and interpretable parameters and the positive definiteness of estimate of Σ_i is guaranteed. In addition, a broad range of covariance structures including stationary and nonstationary structures can be modelled by this approach (see subsection 2.2 and section 5). The more general modelling approach supplements the special structures available for Σ_i in the literature and in some cases subsumes them under the regression umbrella. Moreover, this approach can handle data structure with balance design but subject to monotone dropout and it is easily extendable to multivariate models with more than two response variables and multivariate random coefficient models. Additional development would be required for data sets with intermittent missingness.

The outline of the paper is as follows. §2 introduces the bivariate marginal model we studied and an unconstrained parameterisation and generalized linear regression models for the bivariate covariance matrix. The method of maximum likelihood (ML) is used to estimate the parameters and asymptotic properties of the estimators (of the parameters in mean *and* covariance matrix) are given in §3. In §4 we carry out a simulation study to investigate the asymptotics and an analysis of a bivariate data set is conducted in §5. Finally, in the Appendix we outline the steps underpinning the ML computations and sketch the proofs of theorems in §3.

2. Model Formulation

2·1. *Bivariate Marginal Model*

Consider a balanced repeated measures study in a bivariate case. Let $y_{ij} = (y_{ij}^{(1)}, y_{ij}^{(2)})'$ present the observations of two response variables for the ith individual at jth time point $(i = 1, \cdots, n; j = 1, \cdots, m)$. Assume that y_{ij} arises from the regression model $y_{ij} = x_{ij}\beta + \epsilon_{ij}$, where the $x_{ij} = \begin{pmatrix} x_{ij}^{(1)} & 0 \\ 0 & x_{ij}^{(2)} \end{pmatrix}$ is a $2\times(p_1+p_2+2)$ design matrix with vectors $x_{ij}^{(1)} = (1, x_{ij1}^{(1)}, \cdots, x_{ijp_1}^{(1)})$ and $x_{ij}^{(2)} = (1, x_{ij1}^{(2)}, \cdots, x_{ijp_2}^{(2)})$; Let $p = p_1 + p_2 + 2$; the $\beta = (\beta_{(1)}', \beta_{(2)}')'$ is an unknown $p \times 1$ regression coefficient vector with $\beta_{(1)} = (\beta_0^{(1)}, \beta_1^{(1)}, \cdots, \beta_{p_1}^{(1)})'$ and $\beta_{(2)} = (\beta_0^{(2)}, \beta_1^{(2)}, \cdots, \beta_{p_2}^{(2)})'$; random error $\epsilon_{ij} = (\epsilon_{ij}^{(1)}, \epsilon_{ij}^{(2)})'$ is sampled from a bivariate Gaussian process.

Further let $y_i = (y_{i1}', \cdots, y_{im}')'$ and $X_i = (x_{i1}', \cdots, x_{im}')'$ and $\epsilon_i = (\epsilon_{i1}', \cdots, \epsilon_{im}')'$. Then the matrix form of the model for ith subject is

$$y_i = X_i\beta + \epsilon_i \tag{2.1}$$

for $i = 1, \cdots, n$. Throughout this paper we assume that the covariance matrices of y_i or ϵ_i are homogeneous across subjects and denote them by Σ. However, this constraint can be relaxed in more general schemes.

2·2. *Block triangular factorization of* Σ

The covariance matrix Σ is written as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2m} \\ \vdots & \vdots & & \vdots \\ \Sigma_{m1} & \Sigma_{m2} & \cdots & \Sigma_{mm} \end{pmatrix}.$$

Here $\Sigma_{jk} = E(\epsilon_{ij}\epsilon'_{ik})$ are 2×2 sub-matrices for $j, k = 1, \cdots, m$ and $\Sigma_{jk} = \Sigma'_{kj}$. Noting that Σ is positive definite, Σ can be factorized block-triangularly as (Hamilton 1994)

$$T\Sigma T' = D, \tag{2.2}$$

where T is a block lower triangular with 2×2 identity sub-matrices as diagonal entries and D is a block-diagonal matrix with positive definite 2×2 sub-matrices as diagonal entries. It is easily deduced that the block triangular factorisation of Σ is unique and Σ is positive definite, if and only if, D is positive definite.

Denote D by $D = diag(D_1, \cdots, D_m)$ and T by

$$T = \begin{pmatrix} I & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\Phi_{21} & I & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ -\Phi_{m1} & \cdots & -\Phi_{m(m-1)} & I \end{pmatrix},$$

where I is a 2×2 identity sub-matrix and $\mathbf{0}$ is a 2×2 zero sub-matrix. For simplicity, suppose $\mu_{ij} = E(y_{ij}) = 0$. Let $\widehat{y}_{i1} = 0$ and $\widehat{y}_{ij} = \sum_{k=1}^{j-1} \Phi_{jk} y_{ik}$ for $j = 2, \cdots, m$. Further let

$$\varepsilon_{ij} = y_{ij} - \widehat{y}_{ij} = y_{ij} - \sum_{k=1}^{j-1} \Phi_{jk} y_{ik} \quad \text{for} \quad j = 1, \cdots, m$$

From the standard theory of linear prediction in time series (Hamilton 1994), it can be shown that $\widehat{y}_{ij}$ is the linear least-squares predictor of y_{ij} based on its predecessors $y_{ij-1}, \cdots, y_{i1}$. Thus $\varepsilon_{ij}, j = 1, \cdots, m$ are the prediction errors.

Set $\varepsilon_i = (\varepsilon'_{i1}, \cdots, \varepsilon'_{im})'$, then $\varepsilon_i = Ty_i$. It follows that

$$cov(\varepsilon_i) = Tcov(y_i)T' = T\Sigma T' = D.$$

From above, the block triangular factorization of Σ has the following statistical interpretation: the sub-matrices as the below-diagonal entries of T are the negatives of the coefficient matrices of $\widehat{y}_{ij} = \mu_{ij} + \sum_{k=1}^{j-1} \Phi_{jk}(y_{ik} - \mu_{ik})$, the linear least-squares predictor of y_j based on its predecessors $y_{j-1}, \cdots, y_1$, and block diagonal entries of D are the prediction error covariances $D_j = cov(y_{ij} - \widehat{y}_{ij})$, for $1 \leqslant j \leqslant m$. In addition, the sub-matrices Φ_{jk} in T are unconstrained and block diagonal matrix D is much simpler than Σ. We refer to the new parameters Φ_{jk}'s and D_j's as the generalized autoregressive matrices and innovation covariance matrices of Σ. The block triangular factorization Σ for some special structures such as bivariate compound symmetry, bivariate $AR(1)$ and bivariate $MA(1)$ are investigated (the results can be obtained from the authors). Especially, for the bivariate $AR(1)$, T and D have very simple structures: the first sub-diagonal entries in T are same and the rest of sub-diagonals in T are a zero matrix and the main diagonal entries in D are same except the first entry.

As next, a matrix logarithmic transformation will be applied to D in such way that the resulting estimate must be positive definite and this will underpin our estimation procedure.

2·3. *Matrix logarithm of D*

Suppose D is positive definite, i.e, all the diagonal entries $D_1, \cdots, D_m$ are positive definite. We note the spectral decomposition of, say D_j, is $D_j = C_j G_j C_j'$, where the columns of the 2×2 orthogonal matrix C_j denote the appropriate eigenvectors of D_j, and G_j is a 2×2 diagonal matrix, with diagonal elements equal to the eigenvalues of D_j. The matrix logarithm of D_j can now be defined by

$$A_j = \log D_j = C_j \log(G_j) C_j', \tag{2.3}$$

where $\log(G_j)$ is a diagonal matrix with diagonal elements equal to the logs of the corresponding eigenvalues of D_j. It is obvious that $A_j = \log D_j$ is a symmetric matrix. By the definition of matrix exponential, it follows that $D_j = \exp(A_j)$ and the positive definite of D_j is guaranteed.

The unconstrained matrix A_j (or $\log D_j$), named as log-innovation covariance matrix, is no longer innovation covariance matrix, but it can be claimed that A_j remains some of features associated with the innovation covariance matrix (see Appendix).

With the block triangular factorization of Σ and matrix-logarithmic transformation of D, our transformations are complete. These two transformations define a one-to-one mapping between all of Φ_{jk}'s and $\log D_j$'s and Σ, therefore there is no identifiability problem.

However the number of new parameters in Φ_{jk}'s and $\log D_j$'s is $m(2m + 1)$, which is the same as in Σ, so modelling these new parameters parsimoniously via regression models is now required.

2·4. *Generalized linear regression models for covariance*

The parameters in Φ_{jk} and $\log D_j$ may be modelled in terms of explanatory variables for example, time or time lag. For the purposes of illustrating model development, we confine attention to models of dependence on covariates in linear regression models as is conventional for the mean vectors.

(a) Modelling log-innovation covariance matrices

When developing a linear regression model for $\log D_j (j = 1, \cdots, m)$, it is useful to consider the vectorization

$$a_j = \text{vec}(A_j),$$

where $A_j = \log D_j$. Here a_j is the 3×1 vector consisting of the upper triangular elements of A_j, arranged in some convenient order. We then propose linear models of the form

$$a_j = \text{vec}(A_j) = W_j \lambda, \tag{2.6}$$

for $j = 1, \cdots, m$, where the W_j are specified design matrices with linearly independent columns and $\lambda = (\lambda_1, \cdots, \lambda_d)'$ is a $d \times 1$ vector of unknown parameters. Any model of the form (2.6) can be rearranged in the form

$$A_j = \log D_j = \sum_{l=1}^{d} \lambda_l V_{jl} \tag{2.7}$$

for $j = 1, \cdots, m$, for judicious choices of the 2×2 design matrices V_{jl}. In this case any element of V_{jl} must either equal zero or some elements of W_j.

(b) Modelling generalised autoregressive coefficient matrices

Consider the vectorization for autoregressive coefficient matrices Φ_{jk} for $k < j$ and $j = 2, \cdots, m$. We have

$$b_{jk} = \text{vec}(\Phi_{jk}),$$

where b_{jk} is the 4×1 vector consisting of all the elements of Φ_{jk}. Linear regression models for these vectors are proposed, which are

$$b_{jk} = \text{vec}(\Phi_{jk}) = H_{jk} \gamma \tag{2.8}$$

for $k < j, j = 2, \cdots, m$, where H_{ij} are specified design matrices and $\gamma = (\gamma_1, \cdots, \gamma_q)'$ is an $q \times 1$ vector of unknown parameters. The form (2.8) can be rearranged in the form

$$\Phi_{jk} = \sum_{l=1}^{q} \gamma_l U_{jkl} \tag{2.9}$$

for $k < j, j = 2, \cdots, m$, and the elements of U_{jkl} are chosen from the elements of H_{jk}.

In this broad framework, a wide range of new models becomes available. As a first step only, we illustrate the use of polynomial models as in MacKenzie and Pan (2004) by extending the class $\mathcal{C}^*$ to the bivariate case.

Example: Polynomial models

Since the A_j's are log-innovation covariance matrices at time j, in the most general case, different polynomials can be fitted for the trajectories of the 3 elements of interest over time. That is, for each element in $A_j = \log D_j = \begin{pmatrix} a_{j1} & a_{j2} \\ a_{j2} & a_{j3} \end{pmatrix}$, we may consider that

$$a_{js} = \lambda_0^{(s)} + \lambda_1^{(s)} j + \cdots + \lambda_{d_s}^{(s)} j^{d_s}$$

for $s = 1, 2, 3$, and $d = d_1 + d_2 + d_3 + 3$ is the number of parameters fitted.

Further, noticing that the matrix Φ_{jk} is the coefficients of the response y_k at time point k in the linear prediction of the response y_j at time point j and setting $\Phi_{jk} = \begin{pmatrix} \phi_{jk1} & \phi_{jk3} \\ \phi_{jk2} & \phi_{jk4} \end{pmatrix}$, we may propose that

$$\phi_{jks} = \gamma_0^{(s)} + \gamma_1^{(s)}(j-k) + \cdots + \gamma_{q_s}^{(s)}(j-k)^{q_s}$$

for $s = 1, 2, 3, 4$ and $q = q_1 + q_2 + q_3 + q_4 + 4$ is the number of parameters fitted.

Of course, a key objective will be to find the most parsimonious model and the development of efficient model selection algorithms (Pan and MacKenze, 2003) may be viewed as a priority. In this connection, very little is known, in general, about the behaviour of log-innovation covariance matrices and generalized autoregressive matrices.

3. Maximum likelihood Estimation

3·1. *Estimation of parameters from the joint model*

From the foregoing, the whole of the scientific interest in the joint model lies in the regression parameter vectors β, γ and λ. Their joint estimation can be accomplished using the maximum likelihood approach. The procedures are summarized briefly here - more details are given in the Appendix.

Appealing to the block triangular factorization of Σ in (2.2), the log-likelihood of β, γ and λ, given $y_1, \cdots, y_n$, satisfies

$$2 \log \ell(\beta, \gamma, \lambda | y_1, \cdots, y_n) = -mn \log(2\pi) - n \log |D| - \sum_{i=1}^{n} r_i' T' D^{-1} T r_i, \tag{3.1}$$

where $r_i = y_i - X_i\beta$ and $T = T(\gamma)$ and $D = D(\lambda)$. The maximum likelihood estimator $(\widehat{\beta}', \widehat{\gamma}', \widehat{\lambda}')'$ is obtained by maximizing the function above.

Fixing γ and λ in (3.1) creates a linear estimation equation in β, and the weighted least squares solution of β is

$$\widetilde{\beta} = \{\sum_{i=1}^{n} X_i'\Sigma^{-1}X_i\}^{-1}\sum_{i=1}^{n} X_i'\Sigma^{-1}y_i. \tag{3.2}$$

Secondly, given β and λ, a linear estimation equation in γ can also be obtained by some simple linear algebra. The solution of the score equation with respect to γ is

$$\widetilde{\gamma} = \{\sum_{i=1}^{n} Z_i^{*'} D^{-1} Z_i^{*}\}^{-1}\sum_{i=1}^{n} Z_i^{*'} D^{-1} r_i, \tag{3.3}$$

where $Z_i^* = (r_{i1}^*, \cdots, r_{iq}^*)$ with $r_{il}^* = U_l^* r_i$. Here $U_l^*(l = 1, \cdots, q)$ are the block lower triangular matrices with off-diagonal matrices $U_{jkl}(k < j, j = 2, \cdots, m)$ and zero matrices as diagonal entries.

Let $|D_j|$ be the determinant of D_j, it follows that $\log|D_j| = tr(A_j) = \sum_{l=1}^{d}\lambda_l tr(V_{jl})$ and $D_j^{-1} = \exp(-A_j)$ for $j = 1, \cdots, m$ by the results in Chiu et al. (1996). Let r_{ik} be a 2×1 vector consisting of the $(2k-1)$th and $2k$th elements in r_i and e_{ij} be a 2×1 vector consisting of the $(2j-1)$th and $2j$th elements in Tr_i. It is easily verified that $e_{ij} = r_{ij} - \widehat{r}_{ij}$ with $\widehat{r}_{ij} = \sum_{k=1}^{j-1}\Phi_{jk}r_{ik}$ for $j = 1, \cdots, m$. The log-likelihood function excluding the constant becomes

$$2\log\ell(\beta,\gamma,\lambda|y_1,\cdots,y_n) \sim -mn\sum_{l=1}^{d}\lambda_l tr(\overline{V}_{\cdot l}) - n\sum_{j=1}^{m} tr\{B_j\exp(-A_j)\}, \tag{3.4}$$

where $\overline{V}_{\cdot l} = \sum_{j=1}^{m} V_{jl}/m$ and $B_j = \sum_{i=1}^{n} e_{ij}e_{ij}'/n$.

Applying the directional derivative of the matrix exponential (see Bellman 1970) to the Taylor series expansion of function (3.4) with respect to λ, it can be shown that the lth element of the partial derivative of $\log\ell$ with respect to λ, denoted by $\partial\log\ell(\beta,\gamma,\lambda)/\partial\lambda$, is

$$\frac{2\partial\log\ell(\beta,\gamma,\lambda)}{\partial\lambda_l} = -mn\mathrm{tr}(\overline{V}_{\cdot l}) + n\sum_{j=1}^{m} S(\beta,\gamma,\lambda)_l^j \tag{3.5}$$

for $l = 1, \cdots, d$ and the (l, s)th element of the second-order derivative of $\log\ell$ with respect to λ, denoted by $\partial^2\log\ell(\beta,\gamma,\lambda)/\partial\lambda\partial\lambda'$, is

$$\frac{2\partial^2\log\ell(\beta,\gamma,\lambda)}{\partial\lambda_l\partial\lambda_s} = -n\sum_{j=1}^{m} H(\beta,\gamma,\lambda)_{ls}^j \tag{3.6}$$

for $l, s = 1, \cdots, d$, where

$$S(\beta,\gamma,\lambda)_l^j = \int_0^1 \mathrm{tr}[\exp(-A_jv)B_j\exp\{-A_j(1-v)\}V_{jl}]dv$$

and

$$H(\beta,\gamma,\lambda)^j_{ls} = \int_0^1\int_0^v \mathrm{tr}[\exp(-A_j u)B_j\exp\{-A_j(1-v)\}V_{jl}\times\exp\{-A_j(v-u)\}V_{js}]dudv$$

The integrations in $S(\beta,\gamma,\lambda)^j_l$ and $H(\beta,\gamma,\lambda)^j_{ls}$ can be computed analytically (see Appendix).

Fixed β and γ, the solution of the estimation equation for λ can be obtained by the Newton-Raphson iterations, i.e,

$$\widetilde{\lambda}^{(t+1)} = \widetilde{\lambda}^{(t)} - [\frac{\partial^2 \log\ell(\beta,\gamma,\lambda)}{\partial\lambda\partial\lambda'}]^{-1}\Big|_{\widetilde{\lambda}^{(t)}} \cdot \frac{\partial\log\ell(\beta,\gamma,\lambda)}{\partial\lambda}\Big|_{\widetilde{\lambda}^{(t)}}, \tag{3.7}$$

where $\widetilde{\lambda}^{(t)}$ is the estimate at the tth iteration.

The iterative procedure proceeds within (3.2), (3.3) and (3.7) by initializing at $\Sigma = I_m$ where I_m is a $m\times m$ identity matrix and iterating until convergence to obtain the ML estimator $(\widehat{\beta}', \widehat{\gamma}', \widehat{\lambda}')'$ simultaneously.

3·2. *Asymptotic properties*

In this section the limiting behaviour of the maximum likelihood estimates is presented under some mild regularity conditions. All of the asymptotic results (as $n\to\infty$) take m, p, q and d to be fixed. Let parameter spaces $\mathcal{B}$, Γ and Λ, where $\beta\in\mathcal{B}$, $\gamma\in\Gamma$ and $\lambda\in\Lambda$, be compact subspaces of R^p, R^q and R^d. Let β_0 be the true parameter of β lying in $\mathcal{B}$ and $\alpha_0 = (\gamma_0', \lambda_0')'$ be the true parameter of $\alpha = (\gamma', \lambda')'$ lying in $\Gamma\times\Lambda$. Moreover, for the remainder of the article, it is assumed that

Condition I. For all $\alpha \neq \alpha_0$ in $\Gamma\times\Lambda$,

$$\Sigma(\alpha) \neq \Sigma(\alpha_0),$$

where $\Sigma(\alpha) = T(\gamma)^{-1}D(\lambda)T'(\gamma)^{-1}$. Condition I is needed to guarantee that the density function $f(y_i;\beta,\alpha)$ of Y_i is identifiable. Here Y_i is the random vector for the observation y_i. Given linear regression models of λ, condition I can be always satisfied. The details are given in the simulation section. The consistency of the maximum likelihood estimate is presented in Theorem 1.

Theorem 1. Suppose that the design matrices X_i for $i = 1, 2, \cdots$, are bounded uniformly in the sense that there exists a real number c such that $|(X_i)_{ls}| \le c$ where $(X_i)_{ls}$ is the (l,s)th element of X_i. Moreover suppose that the limit of $\frac{1}{n}\sum_{i=1}^n X_i'\Sigma^{-1}(\alpha)X_i$ exists for $\alpha\in\Gamma\times\Lambda$. Then the maximum likelihood estimate $(\widehat{\beta}', \widehat{\alpha}')'$ is strongly consistent for $(\beta_0', \alpha_0')'$, that is,

$$(\widehat{\beta}', \widehat{\alpha}')' \xrightarrow{a.s} (\beta_0', \alpha_0')' \quad \text{as} \quad n\to\infty.$$

The second theorem establishes the asymptotic normality of the maximum likelihood estimate $(\widehat{\beta}', \widehat{\alpha}')'$.

THEOREM 2. Suppose that

$$V_n^{11}(\beta_0, \alpha_0) = \frac{1}{n}\sum_{i=1}^{n} E_0[\frac{\partial \log f(Y_i;\beta,\alpha)}{\partial \beta}\frac{\partial \log f(Y_i;\beta,\alpha)'}{\partial \beta}]\Big|_{\substack{\beta=\beta_0\\ \alpha=\alpha_0}} \rightarrow V^{11}(\beta_0, \alpha_0);$$

$$V_n^{22}(\beta_0, \alpha_0) = \frac{1}{n}\sum_{i=1}^{n} E_0[\frac{\partial \log f(Y_i;\beta,\alpha)}{\partial \alpha}\frac{\partial \log f(Y_i;\beta,\alpha)'}{\partial \alpha}]\Big|_{\substack{\beta=\beta_0\\ \alpha=\alpha_0}} \rightarrow V^{22}(\beta_0, \alpha_0),$$

where $f(y_i;\beta,\alpha)$ is the density function of Y_i; E_0 denotes the expectation operator with $\beta=\beta_0$ and $\alpha=\alpha_0$. All the matrices $V_n^{11}(\beta_0,\alpha_0)$, $V_n^{22}(\beta_0,\alpha_0)$ and $V^{11}(\beta_0,\alpha_0)$, $V^{22}(\beta_0,\alpha_0)$ are assumed to be positive definite.

Then, under the same assumptions in Theorem 1, the maximum likelihood estimate $(\widehat{\beta}', \widehat{\alpha}')'$ is asymptotically normally distributed as follows:

$$n^{\frac{1}{2}}\begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\alpha}-\alpha_0 \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} V^{11}(\beta_0,\alpha_0)^{-1} & 0 \\ 0 & V^{22}(\beta_0,\alpha_0)^{-1} \end{pmatrix}\right).$$

Theorem 2 implies that $\widehat{\beta}$ and $\widehat{\alpha}$ are asymptotically independent. Further, because $(\widehat{\beta}', \widehat{\alpha}')'$ is a consistent estimate for $(\beta_0', \alpha_0')'$, the asymptotic covariance matrix $V^{11}(\beta_0,\alpha_0)^{-1}$ can be estimated by

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\Sigma^{-1}(\widehat{\alpha})X_i\right)^{-1},$$

where $\Sigma^{-1}(\widehat{\alpha}) = T'(\widehat{\gamma})D^{-1}(\widehat{\lambda})T(\widehat{\gamma})$. And the asymptotic covariance matrix $V^{22}(\beta_0,\alpha_0)^{-1}$ can be approximated by the following matrix

$$\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f(y_i;\beta,\alpha)}{\partial\alpha\partial\alpha'}\Big|_{(\widehat{\beta},\widehat{\alpha})}\right)^{-1}.$$

The proof of Theorems 1 and 2 is sketched in the Appendix.

4. SIMULATION

A small-scale simulation study in this section is conducted in order to investigate the adequacy of the asymptotic results.

A sample of n two-dimensional observations vectors Y_i, $i = 1, \cdots, n$ are generated normally at $m = 11$ time-points. The mean of Y_{ij} in $Y_i = (Y_{i1}, \cdots, Y_{im})$ follows

$$\mu_{ij} = \begin{pmatrix} \beta_1 + \beta_2 t_j \\ \beta_3 + \beta_4 t_j \end{pmatrix}$$

for $t_j = -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$. Thus, the design matrix X_{ij} for the subject i at time-point j is given by

$$X_{ij} = \begin{pmatrix} 1 & t_j & 0 & 0 \\ 0 & 0 & 1 & t_j \end{pmatrix}.$$

The covariance matrix Σ_i of Y_i is constructed by $\Sigma_i = T^{-1}D(T')^{-1}$. The block lower triangular T has 2×2 identity matrices as diagonal entries and the negative of matrices

$$\Phi_{jk} = \begin{pmatrix} \gamma_1 + \gamma_2(t_j - t_k) & \gamma_5 + \gamma_6(t_j - t_k) \\ \gamma_7 + \gamma_8(t_j - t_k) & \gamma_3 + \gamma_4(t_j - t_k) \end{pmatrix},$$

for $j = 2, \cdots, m; k < j$, as off-diagonal entries. The block-diagonal matrix D has positive definite 2×2 matrices $D_j = \exp(A_j)$, $j = 1, \cdots, m$ as diagonal matrices with A_j being

$$A_j = \begin{pmatrix} \lambda_1 + \lambda_2 t_j & \lambda_5 + \lambda_6 t_j \\ \lambda_5 + \lambda_6 t_j & \lambda_3 + \lambda_4 t_j \end{pmatrix}.$$

It is easy to see that (4.1) can be rearranged in the form $\Phi_{jk} = \sum_{l=1}^{8} \gamma_l U_{jkl}$ with $U_{jk1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, U_{jk2} = \begin{pmatrix} j-k & 0 \\ 0 & 0 \end{pmatrix}, U_{jk3} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, U_{jk4} = \begin{pmatrix} 0 & 0 \\ 0 & j-k \end{pmatrix}, U_{jk5} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, U_{jk6} = \begin{pmatrix} 0 & j-k \\ 0 & 0 \end{pmatrix}, U_{jk7} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $U_{jk8} = \begin{pmatrix} 0 & 0 \\ j-k & 0 \end{pmatrix}$. A similar rearrangement can be carried out on A_j in (4.2). Noting that the $(1,1)$-th 2×2 sub-matrix in the covariance matrix $\Sigma_i = T^{-1}D(T')^{-1}$ is D_1 and every element of $\log(D_1)$ is a linear function of t_1, it is easy to see that $\Sigma(\alpha) \neq \Sigma(\alpha_0)$ given that $\lambda \neq \lambda_0$ for at least one value of time point t_1 since a new starting observation time can be chosen for t_1 without affecting analysis of the data. Generally, Condition I can be always satisfied provided that the elements of $\log D_1$ are linear regression models of λ.

Table 1 summarizes the simulation results for $n = 60$ subjects, based on 500 replications made with the same values of the covariates. All the calculations were programmed in R language of version 2.11.1. The function *MatrixExp* adopted in the program is from the R package *msm* contributed by Christopher Jackson. Convergence was considered to be obtained when the differences between the current and previous estimates, i.e, $\| \beta^{(k+1)} - \beta^{(k)} \|$, $\| \gamma^{(k+1)} - \gamma^{(k)} \|$ and $\| \lambda^{(k+1)} - \lambda^{(k)} \|$, are all smaller than a given value, say, 0.00001. Here $\| \mathbf{.} \|$ is the norm of a vector.

In Table 1, the *Average estimate* is the average of the estimated parameters over all simulations and the *True value* is the values used in the model. Notice that these average estimated values demonstrate the strong consistency of the estimates. The *St.Dev* is found from the Hessian matrix given in Theorem and the *Root.MSE* is the square root of sample mean-square-error of the estimates. It is interesting that the values corresponding to β_1, β_3, λ_1, λ_3 and λ_5 are much larger than standard deviation of the other estimates. The *Coverage frequency* reports the percentage of times that the true parameter was located in the corresponding 95% confidence intervals during all simulations. Here 95% confidence interval is calculated by $2 \times (1.96 St.Dev)$. Notice that the coverage frequencies verify the asymptotic normality property of the estimates.

Table 1: Simulation Study Result: based on 500 simulations, sample size = 60, and number of repeated measurements = 11

Parameter	True value	Average estimate	St.Dev	Root.MSE	Coverage frequency
β_1	5.0000	5.0043	0.0869	0.0931	93.8%
β_2	−2.0000	−1.9999	0.0200	0.0226	90.8%
β_3	4.0000	3.9820	0.0863	0.0972	92.6%
β_4	−2.0000	−2.0004	0.0200	0.0231	92.0%
γ_1	0.4000	0.3943	0.0264	0.0274	94.6%
γ_2	−0.0360	−0.0347	0.0068	0.0070	94.0%
γ_3	0.4000	0.3932	0.0264	0.0285	94.0%
γ_4	−0.0360	−0.0342	0.0068	0.0073	93.6%
γ_5	0.2000	0.2020	0.0264	0.0254	96.0%
γ_6	−0.0200	−0.0201	0.0068	0.0064	95.8%
γ_7	0.2000	0.2019	0.0264	0.0264	95.8%
γ_8	−0.0200	−0.0202	0.0068	0.0068	96.2%
λ_1	0.8000	0.7905	0.1170	0.1168	95.0%
λ_2	−0.0640	−0.0648	0.0174	0.0177	95.2%
λ_3	0.8000	0.7954	0.1169	0.1154	96.0%
λ_4	−0.0640	−0.0652	0.0174	0.0167	95.8%
λ_5	0.6000	0.6049	0.0839	0.0829	95.2%
λ_6	−0.0600	−0.0603	0.0124	0.0123	96.2%

5. Bivariate Visual Data: preliminary analysis

A randomized controlled trial of Teletherapy for age-related Macular Degeneration (Hart et al., 2002) carried out in three United-Kingdom based hospital units. The objective is to determine whether teletherapy with 6-mV photons can reduce visual loss in patients with subfoveal choroidal neovascularization in age-related macular degeneration. Two hundred three patients were randomly assigned to radiotherapy or observation and scheduled to visit the clinics at 0, 3, 6, 12, and 24 months. Three visual responses, distance visual acuity, near visual acuity and contrast sensitivity were taken for every patient through out the study. Of 203, 4 were found not to satisfy all study entry criteria and hence excluded from the data analysis. Thus there were 99 patients in the treatment group and 100 in the observation group.

This preliminary analysis focuses on modelling the covariance matrix in bivariate case, so two visual responses, distance visual acuity and near visual acuity, are used and the data only from the treatment group are analyzed. There were around 7% missing data which were imputed by the sample means.

Considering we have not yet developed appropriate model selection method, we fit a saturated model for the mean part, i.e, two different 4th order polynomials for the two visual responses respectively over time. For the variance-covariance part we plot regressograms which are similar to Pourahmadi's (1999) versions, but which are based on the matrix logarithm in order to identify approximate models for autoregressive and innovation parameters.

Table 2 gives maximum likelihood estimates(MLE) of parameters for the mean and the values in brackets are the standard deviation of these estimators.

Table 2: MLE of parameters of the means for Distance Visual Acuity(DVA) and Near Visual Acuity(NVA)

	β_0	β_1	β_2	β_3	β_4
DVA	0.854	0.396	-0.011	0.002	-0.013
	(0.026)	(0.032)	(0.020)	(0.016)	(0.013)
NVA	1.089	0.308	-0.046	0.008	-0.012
	(0.028)	(0.034)	(0.021)	(0.019)	(0.016)

Sample log-innovation variances are plotted in Fig 1 (a), which reveals that the $(1,1)$-th and $(2,2)$-th elements in sample log-innovation covariance matrices may be cubic functions of time t and the $(1,2)$-th and $(2,1)$-th elements are nearly constant over time, so we fit two cubic functions and one constant to log-innovation covariances. The fitted log-innovation covariances are shown in Fig.1.(b). Similarly, the sample generalized autoregressive parameters all indicate that cubic functions of lag $(j-k)$ may be a good choice for fitting all the elements in generalized autoregressive matrices. All the plots of sample and fitted generalized autoregressive parameters are omitted here.

One way to check if our covariance modelling method performs well is to reconstruct variance-covariance matrix and compare the reconstructed variance-covariance matrix with the sample variance-covariance matrix. The reconstruction procedure contains two back-transformations: firstly, estimated innovation matrix $\widehat{D}$ can be obtained by using $D_j = \exp(A_j)$ with $\widehat{A}_j(\widehat{\lambda})$ for $j = 1, \cdots, m$; secondly, estimated variance-covariance matrix $\widehat{\Sigma}$ can be reconstructed by using $\Sigma = T^{-1}DT'^{-1}$ with $\widehat{T}(\widehat{\gamma})$. The estimated correlations and variances for DVA and NVA are shown in below the diagonal and the last row in Table 3.

Looking at the table 3, our method fits the covariance-variance structure very well albeit not perfectly. It is worth to notice that the variances for DVA and NVA slowly

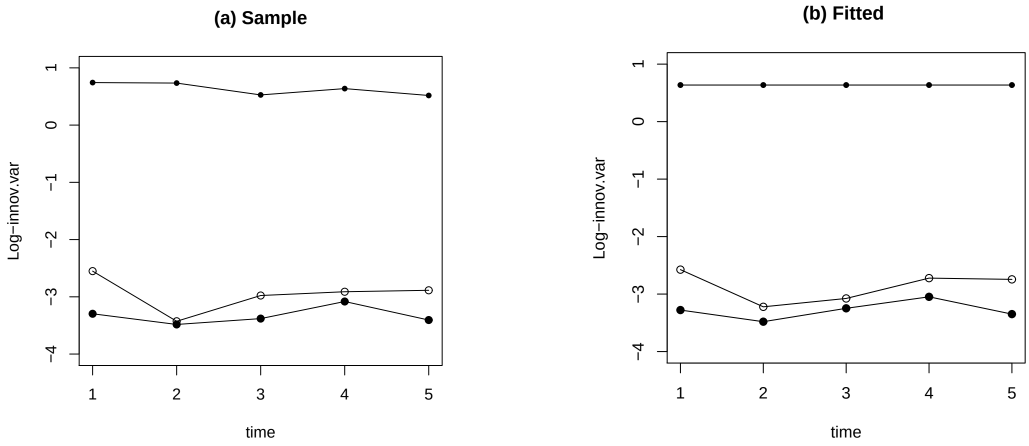

Fig.1. (a) Sample log-innovation variances. joined big dots: the $(1,1)$-th elements; joined circles: the $(2,2)$-th elements; joined small dots: the $(1,2)$-th and $(2,1)$-th elements. (b) fitted log-innovation variances. joined big dots: the fitted cubic polynomial for the $(1,1)$-th elements; joined circles: the fitted cubic polynomial for the $(2,2)$-th elements; joined small dots: the fitted constant for the $(1,2)$-th and $(2,1)$-th elements.

increase over time. This non-stationary feature is well captured by our proposed method. The estimated correlations and cross-correlations between DVA and NVA given in Table 4 also fit the sample correlations and cross-correlations very well (not shown). Inspecting the correlations and cross-correlations in the table reveals at least two patterns: firstly, the correlations are nearly constant except the baseline. Secondly, the correlations in the lower left half seem to be larger than the correlations in the upper right half. This suggests that earlier NVA are more predictive of later DVA, whereas earlier DVA are less predictive of later NVA.

Clearly, the analysis of this set bivariate visual data is preliminary and further detailed analysis is required. For example, not only the mean, but also the covariance structure may depend on the treatment indicator, time and their interaction, so more sophisticated models and treatment effect testing methods might be applicable. The regressograms presented here depend on having balanced data and are not available for unbalanced data sets. In any case, models identified by inspection may not be optimal and new model selection methods such as those proposed by Pan and MacKenzie (2003) for univariate modelling would need to be developed to identify the optimal model in this more complicated joint mean-covariance space.

Table 3: Bivariate visual data. Sample variances (along the main diagonal), sample correlations (above the main diagonal), estimated correlations (below the main diagonal) and estimated variances (last row)

	DVA					NVA				
t	1	2	3	4	5	1	2	3	4	5
1	**0.051**	0.669	0.569	0.542	0.463	**0.096**	0.799	0.636	0.609	0.494
2	0.681	**0.082**	0.835	0.646	0.540	0.678	**0.115**	0.764	0.590	0.504
3	0.466	0.723	**0.138**	0.760	0.598	0.518	0.693	**0.141**	0.685	0.627
4	0.373	0.535	0.744	**0.152**	0.831	0.437	0.530	0.681	**0.142**	0.623
5	0.297	0.455	0.602	0.819	**0.140**	0.302	0.431	0.539	0.720	**0.132**
	0.047	**0.077**	**0.108**	**0.139**	**0.149**	**0.089**	**0.090**	**0.109**	**0.152**	**0.165**

Table 4: Estimated correlation and cross-correlation matrix between DVA and NVA. Estimated correlations (along the main diagonal), estimated cross-correlations at lag k (above the main diagonal), estimated correlations at lag $-k$ (below the main diagonal)

		NVA				
	t	1	2	3	4	5
	1	**0.570**	0.482	0.343	0.229	0.209
	2	0.562	**0.719**	0.554	0.389	0.298
DVA	3	0.503	0.647	**0.748**	0.562	0.425
	4	0.389	0.535	0.653	**0.735**	0.596
	5	0.341	0.459	0.593	0.715	**0.767**

6. Discussion

This paper is a first attempt to develop *data-based* methods for modelling covariance structures in longitudinal studies with multivariate responses. It has direct relevance to the longitudinal randomized clinical trial setting which first stirred the second author's interest in the subject (Hart et al., 2002). It also appears to open several methodological doors, since all of the techniques presented above can, in principle, be extended directly to multivariate case. Broadly speaking, this means that all of the methodological work published in the univariate longitudinal setting can now be extended, relatively straightforwardly, to the multivariate case. Naturally, at the time of writing, there are several open problems. For example, there is a need to develop optimal model selection methods and to extend the current methodology to the linear mixed models setting. There are also other interesting challenges, especially in relation to modelling innovation matrices, whose properties over time are relatively little studied in biostatistical applications. Accordingly, we hope that the methods presented in this paper will stimulate interest in the topic and impact positively on practice.

Acknowledgement

This paper is one of several emanating from the BIO-SI project (07/MI/012) supported by Science Foundation Ireland (SFI), see www.ul,ie/bio-si. Dr Xu is a SFI postdoctoral fellow and Professor MacKenzie is the Principal Investigator.

Appendix

Some properties of log-innovation matrices $\log D_j$*'s*

Let

$$A_j = \begin{pmatrix} a_{j11} & a_{j12} \\ a_{j21} & a_{j22} \end{pmatrix} = \log D_j = C_j \begin{pmatrix} \log d_{j1} & 0 \\ 0 & \log d_{j2} \end{pmatrix} C_j'.$$

We have that

$$a_{j11} + a_{j22} = \log d_{j1} + \log d_{j2} \quad i.e, \quad \mathrm{tr}(A_j) = \log |D_j|$$

and

$$a_{j12} = a_{j21} \leq (a_{j11} + a_{j22})/2 \quad \text{and} \quad a_{j12}^2 = a_{j21}^2 \leq a_{j11} a_{j22}$$

when $\log d_{j1} > 0$ and $\log d_{j2} > 0$. Here d_{j1} and d_{j2} are the eigenvalues of D_j and the columns of C_j are the corresponding normlized eigenvectors.

The proof of which is straightforward by noting that the columns of C_j are the normalized eigenvectors.

These equalities show that the main diagonal elements a_{j11} and a_{j22} act as variances and the off-diagonal elements a_{j12} and a_{j21} act as covariances in A_j, in some sense.

Computations of $S(\beta, \gamma, \lambda)_l^j$ *and* $H(\beta, \gamma, \lambda)_{ls}^j$

The derivative of $2 \log \ell$ with respect to λ can be found by using Taylor-like expansion of the matrix exponential $\exp(-A_j)$ at $-\widehat{A}_j$ in the direction R. Here $\widehat{A}_j = \sum_{l=1}^d \widehat{\lambda}_l V_{jl}$ and R is defined by $-A_j = -\widehat{A}_j + hR$ with h tending to zero. Knowing that $d(trBX) = tr(BdX)$ where B is constant matrix and X is a function matrix and d is a differential operator, we can get

$$\begin{aligned} \mathrm{tr}\{B_j \exp(-A_j)\} &= \mathrm{tr}\{B_j \exp(-\widehat{A}_j)\} - \int_0^1 \mathrm{tr}[B_j \exp\{-\widehat{A}_j(1-v)\}(A_j - \widehat{A}_j)\exp(-\widehat{A}_j v)]dv \\ &+ \int_0^1 \int_0^v \mathrm{tr}[B_j \exp\{-\widehat{A}_j(1-v)\}(A_j - \widehat{A}_j) \\ &\times \exp\{-\widehat{A}_j(v-u)\}(A_j - \widehat{A}_j)\exp(\widehat{A}_j u)]du dv + \text{higher order terms.} \end{aligned}$$

Substituting this equation into (3.4), the Taylor series expansion of $2\log\ell$ with respect to λ is

$$\begin{aligned} 2\log\ell(\beta,\gamma,\lambda|y_1,\cdots,y_n) &= \text{constant} - mn\sum_{l=1}^{d} tr(\overline{V}_{\cdot l})(\lambda_l-\widehat{\lambda}_l) - n\sum_{j=1}^{m} tr\{B_j\exp(-\widehat{A}_j)\} \\ &+ n\sum_{j=1}^{m}\sum_{l=1}^{d} S(\beta,\gamma,\lambda)_l^j(\lambda_l-\widehat{\lambda}_l) \\ &- n\sum_{j=1}^{m}\sum_{l,s=1}^{d} H(\beta,\gamma,\lambda)_{ls}^j(\lambda_l-\widehat{\lambda}_l)(\lambda_s-\widehat{\lambda}_s) + \text{higher order terms,} \end{aligned}$$

where

$$S(\beta,\gamma,\lambda)_l^j = \int_0^1 \text{tr}[\exp(-A_jv)B_j\exp\{-A_j(1-v)\}V_{jl}]dv$$

and

$$H(\beta,\gamma,\lambda)_{ls}^j = \int_0^1\int_0^v \text{tr}[\exp(-A_ju)B_j\exp\{-A_j(1-v)\}V_{jl}\times\exp\{-A_j(v-u)\}V_{js}]dudv.$$

Hence the equation (3.5) and (3.6) hold by uniqueness of Taylor series expansions. The integrations in $S(\beta,\gamma,\lambda)_l^j$ and $H(\beta,\gamma,\lambda)_{ls}^j$ can be computed analytically, using spectral decomposition $\exp(A_j) = C_jG_jC_j'$, where $G_j = \text{diag}(g_{j1}, g_{j2})$ is the diagonal matrix of eigenvalues, and the columns of the C_j are the corresponding normalized eigenvectors. It follows by the definition of matrix exponential that $\exp(-A_jt) = C_jG_j^*C_j'$ with $G_j^* = \text{diag}(g_{j1}^{-t}, g_{j2}^{-t})$. Set $C_j'B_jC_j = \begin{pmatrix} b_{j1} & b_{j3} \\ b_{j3} & b_{j2}\end{pmatrix}$ and $C_j'V_{jl}C_j = \begin{pmatrix} c_{jl1} & c_{jl3} \\ c_{jl4} & c_{jl2}\end{pmatrix}$ and $C_j'V_{js}C_j = \begin{pmatrix} c_{js1} & c_{js3} \\ c_{js4} & c_{js2}\end{pmatrix}$. After some straightforward computation, we can simplify $S(\beta,\gamma,\lambda)_l^j$ and $H(\beta,\gamma,\lambda)_{ls}^j$ as

$$S(\beta,\gamma,\lambda)_l^j = b_{j1}c_{jl1}/g_{j1} + b_{j2}c_{jl2}/g_{j2} - b_{j3}(c_{jl3}+c_{jl4})\log^{-1}(g_{j1}/g_{j2})(1/g_{j1}-1/g_{j2})$$

and

$$\begin{aligned} H(\beta,\gamma,\lambda)_{ls}^j &= \frac{1}{2}\Big(\frac{b_{j1}c_{jl1}c_{js1}}{g_{j1}} + \frac{b_{j2}c_{jl2}c_{js2}}{g_{j2}}\Big) \\ &+ \log^{-1}(\frac{g_{j1}}{g_{j2}})\Big\{b_{j3}(\frac{c_{jl3}c_{js2}}{g_{j2}} - \frac{c_{jl4}c_{js1}}{g_{j1}}) + b_{j3}(\frac{c_{jl2}c_{js4}}{g_{j2}} - \frac{c_{jl1}c_{js3}}{g_{j1}}) \\ &+ (\frac{b_{j2}c_{jl4}c_{js3}}{g_{j2}} - \frac{b_{j1}c_{jl3}c_{js4}}{g_{j1}})\Big\} + \log^{-2}(\frac{g_{j1}}{g_{j2}})(\frac{1}{g_{i1}} - \frac{1}{g_{i2}})\Big\{b_{j3}(c_{jl3}c_{js2} - c_{jl4}c_{js1}) \\ &+ b_{j3}(c_{jl2}c_{js4} - c_{jl1}c_{js3}) + (b_{j2}c_{jl4}c_{js3} - b_{j1}c_{jl3}c_{js4})\Big\}. \end{aligned}$$

Sketch of proof of Theorems Our proof is essentially the same as the proofs of Theorems 1 and 2 in Chiu et al.(1996). Thus, we point out only those differences in computing certain moments that are mostly due to our different reparametrization of Σ for bivariate longitudinal data.

Computation of the moments in Theorem 1

Let $f(y_i;\beta,\gamma,\lambda)$ be the density of Y_i. Then the log pdf of Y_i is

$$\log f(y_i;\beta,\gamma,\lambda) = -\frac{m}{2}\log(2\pi) - \frac{m}{2}\sum_{l=1}^{d}\lambda_l \mathrm{tr}(\overline{V}_{\cdot l}) - \frac{1}{2}(y_i - X_i\beta)'\Sigma^{-1}(\gamma,\lambda)(y_i - X_i\beta)$$

The expectation and the variance of log $f(y_i;\beta,\gamma,\lambda)$ when $\beta=\beta_0$, $\gamma=\gamma_0$ and $\lambda=\lambda_0$, are

$$\begin{aligned} E_0[logf(Y_i;\beta,\gamma,\lambda)] = &- \frac{m}{2}\log(2\pi) - \frac{m}{2}\sum_{l=1}^{d}\lambda_l \mathrm{tr}(\overline{V}_{\cdot l}) - \frac{1}{2}tr\{\Sigma(\gamma_0,\lambda_0)\Sigma^{-1}(\gamma,\lambda)\} \\ &- \frac{1}{2}(\beta_0-\beta)'X_i'\Sigma^{-1}(\gamma,\lambda)X_i(\beta_0-\beta) \end{aligned}$$

and

$$\begin{aligned} V_0[logf(Y_i;\beta,\gamma,\lambda)] &= \frac{1}{2}tr\{\Sigma(\gamma_0,\lambda_0)\Sigma^{-1}(\gamma,\lambda)\}^2 \\ &+ (\beta_0-\beta)'X_i'\Sigma^{-1}(\gamma,\lambda)\Sigma(\gamma_0,\lambda_0)\Sigma^{-1}(\gamma,\lambda)X_i(\beta_0-\beta) \end{aligned}$$

where

$$\Sigma^{-1}(\gamma,\lambda) = T'(\gamma)D^{-1}(\lambda)T(\lambda) = T'(\gamma)\mathrm{diag}\{\exp(-\sum_{l=1}^{d}\lambda_l V_{1l}),\cdots,\exp(-\sum_{l=1}^{d}\lambda_l V_{ml})\}T(\gamma).$$

It follows $V_0[logf(Y_i;\beta,\gamma,\lambda)] \leqslant c_0$ for all i from the compactness of those parameter spaces $\mathcal{B}$, Γ and Λ and boundedness of all the elements in all X_i. Readers are referred to Chiu et al.(1996, p.207) for the proof of Theorem 1.

Computation of $V_n^{11}(\beta_0,\alpha_0)$ and $V_n^{22}(\beta_0,\alpha_0)$ in Theorem 2

Let $\theta = (\beta',\alpha')' = (\beta',\gamma',\lambda')$ and Denote the log-likelihood function of $Y_1,\cdots,Y_n$ by $L(Y_1,\cdots,Y_n;\theta)$, it follows that $L(Y_1,\cdots,Y_n;\theta)$ is regular with respect to its first and second derivatives, i.e.,

$$E\partial L/\partial\theta = 0 \quad \text{and} \quad -E\partial^2 L/\partial\theta\partial\theta' = E\partial L/\partial\theta\partial L/\partial\theta',$$

Denote $\partial L/\partial\beta$, $\partial L/\partial\gamma$ and $\partial L/\partial\lambda$ by $U_1(\beta)$, $U_2(\gamma)$ and $U_3(\gamma)$ respectively. The covariance matrix of the function $n^{-\frac{1}{2}}\partial L/\partial\theta = n^{-\frac{1}{2}}(U_1'(\beta),U_2'(\gamma),U_3'(\lambda))'$ is denoted by $V_n = (v_n^{kl})$ for $k,l = 1,2,3$. Noting the regular condition of the second order derivatives, the covariance matrix V_n can be obtained by computing the expectation of the second order derivatives of $L(Y_1,\cdots,Y_n;\theta)$ at the true value $\theta=\theta_0$. It can also be seen that $V_n^{11}(\beta_0,\alpha_0)$ and $V_n^{22}(\beta_0,\alpha_0)$ in Theorem 2 are equal to the block matrices v_n^{11} and $\begin{pmatrix} v_n^{22} & v_n^{23} \\ v_n^{32} & v_n^{33} \end{pmatrix}$ in V_n respectively.

It is easy to verify that

$$v_n^{11} = -n^{-1}E_0\partial^2 L/\partial\beta\beta'\Big|_{\theta=\theta_0} = n^{-1}\sum_{i=1}^{n}X_i'\Sigma_i^{-1}(\gamma_0,\lambda_0)X_i,$$

where E_0 denotes the expectation operator at the point $\theta=\theta_0$.

The covariance matrix of $n^{-\frac{1}{2}}U_2(\gamma)$ is

$$v_n^{22} = -n^{-1}E_0\partial^2 L/\partial\gamma\gamma'\Big|_{\theta=\theta_0} = n^{-1}E_0\sum_{i=1}^{n} Z_i^{*'}D^{-1}Z_i^*\Big|_{\theta=\theta_0}.$$

Since $Z_i^* = (r_{i1}^*, \cdots, r_{iq}^*)$ with $r_{il}^* = U_l^* r_i (l = 1, \cdots, q)$, it follows that the (l,s)th element of $Z_i^{*'}D^{-1}Z_i^*$ is $r_i' U_l^{*'} D^{-1} U_s^* r_i$ for $l, s = 1, \cdots, q$. Then the (l,s)th element of $E_0\sum_{i=1}^{n} Z_i^{*'}D^{-1}Z_i^*$ is

$$\sum_{i=1}^{n}\Big\{(\beta_0-\beta)'X_i'U_l^{*'}D^{-1}U_s^*X_i(\beta_0-\beta) + \mathrm{tr}(U_l^{*'}D^{-1}U_s^*\Sigma(\gamma_0,\lambda_0))\Big\}.$$

Hence the (l,s)th element of v_n^{22} is

$$\mathrm{tr}(U_l^{*'}D^{-1}(\lambda_0)U_s^*\Sigma(\gamma_0,\lambda_0)).$$

Using the result (3.6) in subsection 3.1, we find that the (l,s)th element in the covariance matrix

$$v_n^{33} = -n^{-1}E_0\partial^2 L/\partial\lambda\lambda'\Big|_{\theta=\theta_0}$$

is

$$E_0\frac{1}{2}\sum_{j=1}^{m} H(\beta,\gamma,\lambda)_{ls}^{j}\Big|_{\theta=\theta_0} =$$
$$\frac{1}{2}\sum_{j=1}^{m}\int_0^1\int_0^v \mathrm{tr}[\exp(-A_{0j}u)D_j(\lambda_0)\exp\{-A_{0j}(1-v)\}V_{jl}\times\exp\{-A_{0j}(v-u)\}V_{js}]dudv,$$

where $\exp\{-A_{0j}\} = D_j^{-1}(\lambda_0)$.

After some matrix algebra, we can see that the factors containing random error in the elements of the matrices $\partial^2 L/\partial\beta\partial\gamma'$ and $\partial^2 L/\partial\beta\partial\lambda'$ are the residual vectors $r_i, i = 1, \cdots, n$, the expectations of which are zero at the true value $\theta = \theta_0$, so it can be obtained that

$$v_n^{12} = 0 \quad \text{and} \quad v_n^{13} = 0.$$

Using the result (3.5) in subsection 3.1, we get that the (l,s)th element in the covariance matrix

$$v_n^{32} = v_n^{23} = -n^{-1}E_0\partial^2 L/\partial\lambda\gamma'\Big|_{\theta=\theta_0}$$

is

$$E_0\Big(-\frac{1}{2}\sum_{j=1}^{m}\frac{\partial S(\beta,\gamma,\lambda)_l^j}{\partial\gamma_s}\Big|_{\theta=\theta_0}\Big) = \frac{1}{2}\sum_{j=1}^{m}\int_0^1 \mathrm{tr}[\exp(-A_{0j}v)B_{js}\exp\{-A_{0j}(1-v)\}V_{jl}]dv$$

Here $B_{js} = \sum_{k=1}^{j-1}(\Sigma_{jk}U_{jks}' + U_{jks}\Sigma_{kj}) - \sum_{k_1=1}^{j-1}\sum_{k_2=1}^{j-1}(\Phi_{jk_1}\Sigma_{k_1k_2}U_{jk_2s}' + U_{jk_1s}\Sigma_{k_1k_2}\Phi_{jk_2}')$.

References

BARNARD, J., MCCULLOCH, R. and MENG, X. (2000). Modelling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage, *Statist. Sinica.* **10**, 1281-1312.

BELLMAN, R. (1970). *Introduction to Matrix Analysis.* New York: McGrawHill.

BENSMAIL, H. and CELEUX, G. (1996) Regularized Gaussian discriminant analysis through eigenvalue decomposition. *J.Am.Statist.Assoc.* **91**, 1743-1748.

BOIK, R.J. (2002). Spectral models for covariance matrices. *Biometrika.* **89**, 159-182.

BOIK, R.J. (2003). Principal component models for correlation matrices. *Biometrika.* **90**, 679-770.

CHEN, Z. and DUNSON, D. (2003). Random effects selection in linear mixed models. *Biometrics.* **59**, 762-769.

CHIU, T.Y.M., LEONARD, T. and TSUI, K.W. (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association*, Vol. **91**, No. 433, 198-210.

FANG, H., TIAN, G., XIONG, X. and TAN, M. (2006). A multivariate random-effects model with restricted parameters: Application to assessing radiation therapy for brain tumours. *Statist. Med***25**: 1948-1959.

FLURY, B. (1984). Common principal components analysis. *J.Amer.Statist.Assoc*, **79**, 892-898.

GALECKI, A.T. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics-Theory and Methods*, **23**: 3105-3119.

GAO, F., THOMPSON, P., XIONG, C.J. and MILLER, J.P. (2006). Analyzing multivariate longitudinal data using SAS. *Proceedings of the Thirty-first Annual SAS Uers Group International Conference.* Cary, NC: SAS Institute Inc. Paper 187-31.

HAMILTON, J. D.(1994). *Time Series Analysis.* Princeton, New Jersey.

HART, P. M., CHAKRAVARTHY, U., MACKENZIE, G., CHISHOLM, I. H., BIRD, A. C., STEVENSON, M. R., OWENS, S. L., HALL, V., HOUSTON, R. F., MCCULLOCH, D. W. & PLOWMAN, N. (2002). Visual outcomes in the subfoveal radiotherapy study. *Archives of Ophthalmology* **120**, 1029-1038.

JONES, R.H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach.* London: Chapman and Hall.

MACKENZIE, G. and PAN, JX (2004). Optimal joint-mean covariance modelling. *Proceedings of the 2nd International Workshop on Correlated Data Modelling.* Editors: MacKenzie G. and Gregori D. (Pub:July, 2007,
Franco Angeli, Milan. http://www.staff.ul.ie/mackenzieg/WCDM/wcdm.html)

POURAHMADI, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**: 677-690.

PAN, J.X. and MACKENZIE, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika* **90**, 239-244.

PINHIERO, J. and BATES, D.M. (1996). Unconstrained parameterisations for variance-covariance matrices. *Statist. Comp.* **6**, 289-296.

REEVES, J. and MACKENZIE, G. (1998). A bivariate regression model with serial correlation.

The Statistician, Vol. **47**, No. 4, 607-615.

SMITH, M. and JOHN, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *J. Amer. Statist. Assoc.* **97**, 1141-1153.