Modelling Conditional Covariance Structure in Linear Mixed Models

Jianxin Pan¹, Gilbert MacKenzie¹

¹ The Centre for Medical Statistics, Keele University, Keele, Staffordshire, ST5 5BG, UK. Email: j.x.pan@keele.ac.uk; g.mackenzie@keele.ac.uk

Keywords: Covariance Structure, Linear Mixed Models, Polynomial Fitting, Maximum Hierarchical Likelihood Estimation.

1 Introduction

Linear mixed models (LMMs) are widely used for analysis of longitudinal data because correlation in repeated measures over time for each subject is taken into account by incorporation of random effects (e.g., Diggle, Liang and Zeger, 1994). The general form of LMMs can be written as

\[ Y_i = X_i \beta + Z_i u_i + \epsilon_i, \]

where, for the \(i\)th subject, \(Y_i\) is the \((m_i \times 1)\) stacked vector of \(m_i\) responses made over time, \(X_i\) is a \((m_i \times p)\) matrix of covariates, \(\beta\) is a \((p \times 1)\) vector of unknown fixed effects, \(Z_i\) is a \((m_i \times q)\) design matrix for the \((q \times 1)\) vector of between subjects random effects \(u_i\), and \(\epsilon_i\) is a \((m_i \times 1)\) vector of residuals, for \(i = 1, \ldots, n\) subjects. It is usual to adopt a two-stage hierarchical modelling approach in which \(u_i \sim N(0, G)\) and \(\epsilon_i | u_i \sim N(0, R_i)\) where \(G\) is the \((q \times q)\) between subjects covariance matrix, constant for all subjects, and \(R_i\) is the \((m_i \times m_i)\) conditional covariance matrix for the repeated measurements made on the \(i\)th subject, given the random effects \(u_i\).

These arrangements lead, after integrating out \(u_i\), to a marginal model in which \(E(Y_i) = X_i \beta\) and \(V(Y_i) = Z_i G Z_i' + R_i = \Sigma_i\). In longitudinal data analysis, the usual parametric specification for \(R_i\) and \(G\) is assuming \(R_i = \sigma^2_i I_{m_i}\) and \(G = \sigma^2_u I_q\). In other words, the repeated measures over time are conditionally independent and the components of the random effects \(u_i\) are independent as well. In this case, the marginal covariance matrix \(\Sigma_i\) has the property of compound symmetry. However, the assumption of conditional independence may be unreasonable in practice since, when it holds, the correlation between measurements on the same subject is generated principally by the magnitude of the between-subject variation (Reeves and MacKenzie, 1998). Accordingly, it is usual to adopt a
particular within subject covariance structure for $R_i$ - AR(1), AR(2) and unstructured covariance models are frequently considered in the literature. Alternatively, a Gaussian stochastic process (Diggle, Liang and Zeger, 1994) may be adopted.

However, such methods are _menu-based_ and problems may arise when the selected covariance structure is very different from the true structure. For example, the mis-specification may bias the estimate of the fixed effects in finite samples. Pourahmadi (1999) proposed a more flexible _data-driven_ approach in which any _marginal_ covariance matrix arising in longitudinal studies may be modelled using a polynomial of time.

In longitudinal studies, however, the assumption of a homogeneous marginal covariance structure is a _testable_ model choice and Pan and MacKenzie (2000) generalised Pourahmadi's approach by including baseline covariates in the specification of the marginal covariance matrix arising in LMMs, i.e., from $\Sigma(t; \theta) \rightarrow \Sigma(t, \theta, x, \beta^*)$.

In the framework of LMMs, the parameters in the between subject covariance matrix $G$ are constant across subjects and so any heterogeneity in the marginal covariance should arise as a consequence of heterogeneity in the conditional covariance matrices $R_i$. In this paper, we provide a data-driven approach to detect and explain heterogeneity in the conditional covariance matrices $R_i$. We model the $R_i$ parsimoniously using a regression approach and estimate the parameters by a maximum hierarchical likelihood estimation (MHLE) procedure (Lee and Nelder, 1996) which exploits the information in the estimated marginal covariance matrix. We compare our proposed procedure with standard menu selection methods, reporting results from a simulation study and the analysis of Kenward's cattle data.

### 2 Hierarchical Maximum Likelihood Estimation

Denote $G = G(\theta)$ and $R_i = R_i(\alpha)$ where $\theta$ and $\alpha$ are parameters in $G$ and $R_i$. We use the MHLE procedure to estimate the fixed effects $\beta$ and variance components $\theta$ and $\alpha$, and to predict the random effects $u_i$.

#### 2.1 Estimation of $\alpha$

First, for the conditional covariance matrix $R_i$, there is a unique lower triangular matrix $T_i$ with 1’s as diagonal entries and a unique diagonal matrix $D_i$ with positive diagonal entries such that $T_iR_iT_i' = D_i$. This decomposition has a simple statistical interpretation: given the random effects $u_i$, the below-diagonal entries of $T_i$ are the negatives of the autoregressive coefficients, $\phi_{ijk}$, in $\hat{y}_{ij} = \mu_{ij} + \sum_{k=1}^{j-1} \phi_{ijk}(y_{ik} - \mu_k)$, the linear least squares predictor of $y_{ij}$ based on its predecessors $y_{i(j-1)}$, ..., $y_{i1}$, where $\mu_{ij} = E(y_{ij}|u_i)$. It may be shown that the diagonal entries of $D_i$ are the conditional prediction error (innovation) variances $\sigma_{ij}^2 = \text{var}((y_{ij} - \hat{y}_{ij})|u_i)$. 

where $1 \leq j \leq m_i$ and $1 \leq i \leq n$ (Pourahmadi, 1999). It follows immediately that $T_i^{-1} = T_i' D_i^{-1} T_i'$. Second, we fit the unconstrained parameters $\phi_{ijk}$ and $\log \sigma_j^2$ using two regression models $\phi_{ijk} = a_{ijk}' \gamma$ and $\log \sigma_j^2 = b_{ij}' \lambda$ where $\gamma$ and $\lambda$ are the parsimonious parameters for modelling $R_i$ and $\alpha = (\gamma', \lambda')'$. The covariates vectors $a_{ijk}$ and $b_{ij}$ may contain baseline covariates, polynomial terms of time, and their interactions (Pan and MacKenzie, 2000). Based on the joint likelihood of the response $Y$ and random effects $u$, the parameter $\gamma$ can be estimated by

$$
\hat{\gamma} = \left( \sum_{i=1}^{n} A_i' D_i^{-1} A_i \right)^{-1} \left( \sum_{i=1}^{n} A_i' D_i^{-1} r_i \right)
$$

(2)

provided that the parameters $\beta$ and $\lambda$ and the random effects $u_i$ are given, where $A_i = (a_{i1}, a_{i2}, ..., a_{ilm})'$ with $a_{ijk} = \sum_{k=1}^{j-1} a_{ijk} r_{ik}$ and $r_{ik}$ is the $k$th component of $r_i = Y_i - X_i \beta - Z_i u_i$. Similarly, the one-step updated estimate of $\lambda$ is

$$
\hat{\lambda} = \left( \sum_{i=1}^{n} B_i' V_i^{-1} B_i \right)^{-1} \left( \sum_{i=1}^{n} B_i' V_i^{-1} e_i \right)
$$

(3)

where $B_i = (b_{i1}, b_{i2}, ..., b_{im})'$, $V_i = \text{diag}(v_{i1}^2, v_{i2}^2, ..., v_{im}^2)$ with $v_{ij}^2 = \sigma_j^2 (r_{ij} - \hat{r}_{ij})^{-2}$ and $\hat{r}_{ij} = \sum_{k=1}^{j-1} \phi_{ijk} r_{ik}$, and $e_i = (\log D_i + I_{m_i} - V_i)^1_{m_i}$. Equations (a) and (3) give the MHLE $\hat{\alpha} = (\hat{\gamma}', \hat{\lambda}')'$ of $\alpha$ when the fixed effects $\beta$ and the random effects $u_i$ are given.

### 2.2 Estimation of $\theta$

The MHLE of $\theta$ in $G(\theta)$ maximises the joint likelihood of $Y$ and $u$, or equivalently, maximises $l^* = -(n/2) \log |G| - (1/2) \sum_{i=1}^{n} v_i' G^{-1} u_i$ with respect to $\theta$, provided other parameters are fixed. In general, the solution is not of closed form and one must proceed iteratively. However, analytical solutions exist in special cases. For example, a common structure for $G$ in longitudinal studies is $G(\theta) = \text{diag}(\theta_1 I_{q_1}, \theta_2 I_{q_2}, ..., \theta_t I_{q_t})$, where $\theta = (\theta_1, \theta_2, ..., \theta_t)'$. In this case, the MHLE of $\theta_s$ can be written as

$$
\hat{\theta}_s = \frac{1}{q_s} \sum_{i=1}^{n} \bar{u}_{is} u_{is}
$$

(4)

where $u_{is}$ are sub-blocks of the random effects $u_i$ for $s = 1, 2, ..., c$.

### 2.3 Estimation of $\beta$ and prediction of $u_i$

Given the marginal matrices $\Sigma_i = Z_i G Z_i' + R_i$, maximising the joint likelihood of the response $Y$ and the random effects $u$ with respect to $\beta$ leads to

$$
\hat{\beta} = \left( \sum_{i=1}^{n} X_i' \Sigma_i^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{n} X_i' \Sigma_i^{-1} Y_i \right)
$$

(5)
which is the MHLE of the fixed effects $\beta$. When $\hat{\beta}$ in (5) becomes available, the prediction of the random effects $u_i$ must be of the form

$$\hat{u}_i = GZ_i\Sigma^{-1}_i(Y_i - X_i\hat{\beta}).$$

In summary, given starting values of $G$ and $R_i$, e.g., $G = I_q$ and $R_i = I_{m_i}$, the estimate of $\beta$ and the prediction of $u_i$ can be calculated using (5) and (6). Based on these estimates, the estimates of $\alpha = (\gamma', \lambda')'$ and $\theta$ are then calculated using (2)-(4). We iterate this procedure until convergence and the HMLEs of parameters are thus obtained.

3 Simulation Studies and Examples

To compare the proposed data-driven approach with menu-based methods, we conduct a simulation study in which the true structure of the conditional covariance is generated by an exponential function of time and it also depends on a specific covariate. We fit the simulated data sets by the use of LMMs with compound symmetry structure, AR(1) structure on the conditional covariance, and the data-driven conditional covariance structure. The simulation study shows that the proposed data-driven approach is capable of detecting heterogeneity in the conditional covariance matrices, capturing the true structure of the covariance matrices, and yielding an accurate estimate of the fixed effects. In contrast, the models with compound symmetry and AR(1) conditional covariance matrices produce a biased estimate of the fixed effects due to the mis-specification for the covariance structure. Kenward’s cattle data (Pourahmadi, 1999) are also analysed by the proposed approach. We will provide the details in the full paper if accepted for presentation at the IWSM.

References


