

# Modelling Marginal Covariance Structures in Linear Mixed Models

Gilbert MacKenzie<sup>1</sup> and Jianxin Pan<sup>1</sup>

<sup>1</sup> Centre for Medical Statistics, Keele University, Keele, Staffordshire ST5 5BG, UK. E-mail: g.mackenzie@keele.ac.uk

## Abstract

Pourahmadi (1999) provided a convenient reparameterisation of the marginal covariance matrix arising in longitudinal studies. We exploit his work to model the dependence of this covariance structure on baseline covariates, time and their interaction. The rationale for this approach is the realisation that in linear mixed models (LMMs) the assumption of a homogeneous covariance structure with respect to the covariate space is a *testable* model choice. Accordingly, we provide methods for testing this assumption and re-analyse Kenward's (1987) cattle data set using our new model.

Keywords: Cholesky decomposition; Covariate dependent covariance; Linear mixed models; Longitudinal data

## 1 Introduction

In longitudinal studies the classical linear mixed model of Laird & Ware (1982) is frequently adopted when the response is a set of continuous measurements made over time. In general:

$$Y_i = X_i\beta + Z_iu_i + \epsilon_i \quad (1)$$

where, for the  $i$ th subject,  $Y_i$  is the  $(m_i \times 1)$  stacked vector of  $m_i$  responses made typically at times  $t_0, t_1, \dots, t_{m_i-1}$ ,  $X_i$  is a  $(m_i \times p)$  matrix of covariates some of which may be functions of time,  $\beta$  is a  $(p \times 1)$  vector of unknown regression coefficients,  $Z_i$  is a  $(m_i \times r)$  design matrix for the  $(r \times 1)$  vector of between subjects random effects,  $u_i$ , and  $\epsilon_i$  is a  $(m_i \times 1)$  vector of residuals, for  $i = 1, \dots, n$  subjects and  $m = \sum_{i=1}^n m_i$ . These arrangements lead, after integrating out  $u_i$ , to a marginal model with covariance matrix

$$\Sigma = \Sigma(t; \theta) \quad (2)$$

where  $\theta$  is a low dimensional parameter vector characterising how the covariance structure depends on time.

In many biological and medical problems the variability over time may be influenced by the baseline covariate profile. For example, the effect of intervention in a longitudinal randomised controlled trial may be to modify the covariance structure *per se*, rather than, or perhaps as well as, the marginal mean. We address this issue by extending Pourahmadi's data-driven methods to include baseline covariates in the specification of *any* marginal covariance matrix arising in the general linear model. Effectively, our joint mean-covariance model becomes:

$$\begin{aligned}\mu &= X\beta \\ \Sigma &= \Sigma(t, \theta, x, \beta^*)\end{aligned}\quad (3)$$

where  $\beta^*$  measures the influence of the baseline covariates and their interaction with time on  $\Sigma$ . As we demonstrate below, (3) provides a powerful and convenient framework for estimation and hypothesis testing in longitudinal studies.

## 2 Covariance Regression Model Formulation

### 2.1 Modified Cholesky Decomposition

For the subject-specific covariance matrix  $\Sigma_i$ , there is a unique lower triangular matrix  $T_i$  with 1's as diagonal entries and a unique diagonal matrix  $D_i$  with positive entries such that  $T_i \Sigma_i T_i' = D_i$ . This decomposition has a simple statistical interpretation: the below-diagonal entries of  $T_i$  are the negatives of the autoregressive coefficients,  $\phi_{ijk}$ , and the diagonal entries of  $D_i$  are the prediction error (innovation) variances  $\sigma_{ij}^2 = \text{var}(y_{ij} - \hat{y}_{ij})$  where  $1 \leq j \leq m_i$  and  $1 \leq i \leq n$  (Pourahmadi, 1999). It follows immediately that  $\Sigma_i^{-1} = T_i' D_i^{-1} T_i$ .

### 2.2 Regression Model

The parameters  $\phi_{ijk}$  and  $\varsigma_{ij} = \log \sigma_{ij}^2$  are unconstrained and were modelled as polynomials of time by Pourahmadi (1999). By including covariates our generalised model takes the form;

$$\begin{aligned}\mu_{ij} &= x'_{ij} \beta_1 \\ \phi_{ijk} &= \tilde{x}'_{ijk} \beta_2 + z'_{ijk} \gamma \\ \varsigma_{ij} &= \tilde{x}'_{ij} \beta_3 + z'_{ij} \lambda\end{aligned}\quad (4)$$

a structure depending on five sets of parameters. In (4)  $x_{ij}$  is a  $(p_1 \times 1)$  vector containing the baseline covariates and an intercept term for the mean structure, in which some elements may be functions of time. The  $(p_2 \times 1)$  vector  $\tilde{x}_{ijk}$  and the  $(p_3 \times 1)$  vector  $\tilde{x}_{ij}$  contain baseline covariates for the  $i$ th subject and interaction terms with  $z_{ijk}$  ( $q \times 1$ ) and  $z_{ij}$  ( $d \times 1$ ), respectively. The covariate vectors  $z_{ij}$  and  $z_{ijk}$  are completely determined by the times  $t_{ij}$  and  $t_{ik}$  and the intercept terms.

### 3 Maximum Likelihood Estimation

#### 3.1 The Likelihood

Corresponding to (4) the log-likelihood function has three representations:

$$\begin{aligned}
-2\ell &= \sum_{i=1}^n \log |\Sigma_i| + \sum_{i=1}^n (Y_i - X_i \beta_1)' \Sigma_i^{-1} (Y_i - X_i \beta_1), \\
&= \sum_{i=1}^n \sum_{j=1}^{m_i} \log \sigma_{ij}^2 + \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{(r_{ij} - \hat{r}_{ij})^2}{\sigma_{ij}^2} \\
&= \sum_{i=1}^n \log |D_i| + \sum_{i=1}^n \{r_i - \tilde{X}_i^* \beta_2 - Z_i^* \gamma\}' D_i^{-1} \{r_i - \tilde{X}_i^* \beta_2 - Z_i^* \gamma\}
\end{aligned} \tag{5}$$

where the scalar  $r_{ij} = y_{ij} - x'_{ij} \beta_1$  is the  $j$ th element of  $r_i = Y_i - X_i \beta_1$  the  $(m_i \times 1)$  vector of residuals and the  $(1 \times p_1)$  row vector  $x'_{ij}$  contains baseline covariates propagated through the  $m_i$  time points:  $i = 1, \dots, n$  and  $j = 1, \dots, m_i$ . The scalar  $\hat{r}_{ij} = \sum_{k=1}^{j-1} \phi_{ijk} r_{ik}$  is the predicted residual based on its predecessors and is the  $j$ th element of the vector  $\hat{r}_i$  of order  $(m_i \times 1)$ . The matrix  $\tilde{X}_i^*$ , of order  $(m_i \times p_2)$ , has typical row  $\tilde{x}'_{ij}$  and it may be shown that  $\tilde{x}'_{ij} = \sum_{k=1}^{j-1} r_{ik} \tilde{x}_{ijk}$ , a vector addition indexed by  $k$ . The matrix  $Z_i^*$  of order  $(m_i \times q)$  has a similar structure,  $z'_{ij} = \sum_{k=1}^{j-1} r_{ik} z_{ijk}$ . These representations facilitate the computation of the maximum likelihood estimates (MLEs).

#### 3.2 Maximum Likelihood Estimation

Based on the three representations, the score functions corresponding to the parameters  $\beta_1$ ,  $\beta_2$ ,  $\gamma$ ,  $\beta_3$  and  $\lambda$  are given by:

$$\begin{aligned}
U_1(\beta_1) &= \sum_{i=1}^n X_i' \Sigma_i^{-1} (Y_i - X_i \beta_1) \\
U_2(\beta_2) &= \sum_{i=1}^n \tilde{X}_i^{*'} D_i^{-1} \{r_i - \tilde{X}_i^* \beta_2 - Z_i^* \gamma\} \\
U_3(\gamma) &= \sum_{i=1}^n Z_i^{*'} D_i^{-1} \{r_i - \tilde{X}_i^* \beta_2 - Z_i^* \gamma\} \\
U_4(\beta_3) &= \frac{1}{2} \sum_{i=1}^n \tilde{X}_i' (D_i^{-1} e_i - \mathbf{1}_{n_i}) \\
U_5(\lambda) &= \frac{1}{2} \sum_{i=1}^n Z_i' (D_i^{-1} e_i - \mathbf{1}_{n_i})
\end{aligned} \tag{6}$$

where  $\tilde{X}_i^*$  and  $Z_i^*$  are defined above and  $e_i = (e_{i1}, e_{i2}, \dots, e_{im_i})'$  with  $e_{ij} = (r_{ij} - \hat{r}_{ij})^2$ ,  $\tilde{X}_i = (\tilde{x}'_{i1}, \tilde{x}'_{i2}, \dots, \tilde{x}'_{im_i})'$  and  $Z_i = (z'_{i1}, z'_{i2}, \dots, z'_{im_i})'$ , where  $1_{m_i}$  is the  $m_i \times 1$  vector of 1's.

We calculate the Hessian matrix and the Fisher information matrix as follows. First, the derivative of  $U_1(\beta_1)$  with respect to  $\beta_1$ , i.e.,  $H_{11}$ , takes the form

$$H_{11} = - \sum_{i=1}^m X_i' \Sigma^{-1} X_i \quad \text{so that} \quad I_{11} = \sum_{i=1}^m X_i' \Sigma^{-1} X_i \quad (7)$$

Second, let  $\dot{\Sigma}_{i,\alpha}^{-1}$  be the matrix derivative of  $\Sigma_i^{-1}$  with respect to the vector  $\alpha$ , i.e.,  $\dot{\Sigma}_{i,\alpha}^{-1} = \partial[\text{vec}(\Sigma_i^{-1})]' / \partial \alpha$  (see, e.g., Pan, Fang & von Rosen, 1997), where  $\text{vec}(A)$  is the vector obtained by vectorizing the matrix  $A$  through column by column. By taking derivatives of  $U_1(\beta_1)$  with respect to  $\alpha$ , we obtain

$$\frac{\partial U_1'(\beta_1)}{\partial \alpha} = \sum_{i=1}^m \dot{\Sigma}_{i,\alpha}^{-1} (r_i \otimes X_i) \quad (8)$$

and therefore the expectation of (6) must be zero because of  $E(r_i) = 0$ . In other words, we have  $I_{i1} = 0$  ( $i = 2, 3, 4, 5$ ). Third, after some algebra other block components in the negative of the Hessian matrix, i.e.,  $(-H_{ij})$  are of the form

$$\left( \begin{array}{cccc} \sum_{i=1}^n \tilde{X}_i^{*'} D_i^{-1} \tilde{X}_i^* & & & \\ \sum_{i=1}^n Z_i^{*'} D_i^{-1} \tilde{X}_i^* & \sum_{i=1}^n Z_i^{*'} D_i^{-1} Z_i^* & & \\ \sum_{i=1}^n \tilde{X}_i' R_i^{-1} \tilde{X}_i^* & \sum_{i=1}^n \tilde{X}_i' R_i^{-1} Z_i^* & \frac{1}{2} \sum_{i=1}^n \tilde{X}_i' V_i^{-1} \tilde{X}_i & \\ \sum_{i=1}^n Z_i' R_i^{-1} \tilde{X}_i^* & \sum_{i=1}^n Z_i' R_i^{-1} Z_i^* & \frac{1}{2} \sum_{i=1}^n Z_i' V_i^{-1} \tilde{X}_i & \frac{1}{2} \sum_{i=1}^n Z_i' V_i^{-1} Z_i \end{array} \right) \quad (9)$$

where both  $R_i^{-1} = \text{diag}((r_{i1} - \hat{r}_{i1})/\sigma_{i1}^2, (r_{i2} - \hat{r}_{i2})/\sigma_{i2}^2, \dots, (r_{im_i} - \hat{r}_{im_i})/\sigma_{im_i}^2)$  and  $V_i^{-1} = \text{diag}(e_{i1}/\sigma_{i1}^2, e_{i2}/\sigma_{i2}^2, \dots, e_{im_i}/\sigma_{im_i}^2)$  are diagonal matrices of size  $m_i \times m_i$ . On the other hand, by taking the expected value of (9) one may obtain the remaining block components, i.e.,  $I_{\alpha} = (I_{ij})$  for  $i, j = 2, \dots, 5$  relatively easily. Based on (6) and (9), the MLEs of  $\beta_1, \beta_2, \gamma, \beta_3$  and  $\lambda$  can be calculated using the Newton-Raphson algorithm or the Fisher scoring algorithm.

## 4 An Example

Kenward (1987) analysed an experiment in which 60 cattle were assigned randomly to two treatment groups A and B, and their weights were recorded to study the effect of treatment on intestinal parasites. The animals were weighted 11 times over the 133-day period at 0, 14, 28, 42, 56, 70, 84, 98, 112, 126 and 133 days on schedule. Pourahmadi (2000) analysed the data in treatment group A using a saturated mean model (11 parameters) and from the sample regressograms estimated two cubic polynomials of time to model the covariance structure.

We analyse the whole cattle data set using a saturated mean for the data. However, in addition to cubic polynomials of time, we also take the effects of the baseline treatment indicator,  $a_i$ , into account choosing:

$$\begin{aligned}\tilde{x}_{ij} &= (a_i, a_i * t_{ij}, a_i * t_{ij}^2, a_i * t_{ij}^3)' \\ \tilde{x}_{ijk} &= (a_i * (t_{ij} - t_{ik}), a_i * (t_{ij} - t_{ik})^2, a_i * (t_{ij} - t_{ik})^3)'\end{aligned}$$

where  $a_i = 0$  if the  $i$ th animal received treatment A and one otherwise. The time-dependent covariates  $z_{ij}$  and  $z_{ijk}$  take the simple forms:

$$\begin{aligned}z_{ij} &= (1, t_{ij}, t_{ij}^2, t_{ij}^3)' \\ z_{ijk} &= (1, (t_{ij} - t_{ik}), (t_{ij} - t_{ik})^2, (t_{ij} - t_{ik})^3)'\end{aligned}$$

where  $i = 1, 2, \dots, 60$ ,  $j = 1, 2, \dots, 11$  and  $k = 1, 2, \dots, (j - 1)$ .

We fitted the model using an inter-dependent updating profile likelihood method which we encoded in S-Plus. The algorithm provides closed form, GLS-type, solutions for the parameters of interest at each major stage and converges rapidly to the ML solution. The estimating equation for the innovation variances requires a Newton-Raphson updating algorithm at each major stage. However, this sub-loop requires only one or two iterations and again the update takes the form of a GLS solution. In the example analysed, only one iteration was used in the sub-loop and the algorithm converged. The results were checked using the derivative-free minimiser in S-Plus (V4.5) and also the Fisher scoring algorithm in §3. All the solutions agreed. Further details of the computational algorithm may found in Pan & MacKenzie (2000)

We fitted 4 models of interest. Model I is the *full* model in the sense that covariates and orthogonal polynomials of time are used to model the correlation parameters and innovation variances. In contrast, Models II and III are *partial* models in that covariates are omitted either from the correlation parameters (Model II) or from innovation variances (Model III), while Model IV merely includes orthogonal polynomials of time (Pourahmadi's

model), but applied to the entire data set rather than to treatment group A. The MLEs for the parameters in Model I - IV are summarised in the Table 1 below.

Table 1. *MLEs of the  $(4 \times 1)$  parameter vectors  $\beta_2$  and  $\gamma$  in the correlation parameters and the  $(4 \times 1)$  parameter vectors  $\beta_3$  and  $\lambda$  in the innovation variances for Models I - IV of the marginal covariance matrix derived from the cattle data (estimated standard errors in parentheses)*

Model	Par.	(1)	(2)	(3)	(4)
I	$\hat{\beta}_2'$	0.003 (.01)	- 0.16 (.97)	0.19 (1.43)	-1.72 (1.91)
	$\hat{\gamma}'$	0.181 (.01)	-12.25 (.55)	10.85 (.88)	-7.29 (1.71)
	$\hat{\beta}_3'$	-0.024 (.11)	11.52 (2.83)	11.25 (2.78)	2.65 (2.80)
	$\hat{\lambda}'$	3.672 (.08)	- 4.12 (2.01)	5.03 (1.93)	-5.66 (1.97)
II	$\hat{\gamma}'$	0.182 (.01)	-12.32 (.42)	11.00 (.66)	-7.96 (.90)
	$\hat{\beta}_3'$	-0.022 (.11)	11.27 (2.83)	11.29 (2.78)	2.90 (2.80)
	$\hat{\lambda}'$	3.673 (.08)	- 3.99 (2.00)	5.07 (1.93)	-5.66 (1.96)
III	$\hat{\beta}_2'$	0.004 (.08)	- 0.22 (.96)	0.52 (1.44)	-1.10 (1.96)
	$\hat{\gamma}'$	0.181 (.01)	-12.41 (.66)	10.78 (.97)	-7.77 (1.32)
	$\hat{\lambda}'$	3.709 (.06)	- 2.85 (1.41)	11.39 (1.38)	-4.05 (1.40)
IV	$\hat{\gamma}'$	0.183 (.01)	-12.31 (.48)	11.01 (.72)	-8.26 (.97)
	$\hat{\lambda}'$	3.710 (.06)	- 2.88 (1.41)	11.46 (1.38)	-3.95 (1.40)

Letting  $\hat{\ell}_I$ ,  $\hat{\ell}_{II}$ ,  $\hat{\ell}_{III}$  and  $\hat{\ell}_{IV}$ , represent the maximised log-likelihood functions under Models I - IV, respectively, we may test the null hypothesis of no covariate effect,  $\beta_2 = 0$  and  $\beta_3 = 0$ , using  $-2(\hat{\ell}_{IV} - \hat{\ell}_I) \sim asym \chi_8^2 = 34$ , a highly significant result, the critical value for  $\chi^2$  being 15.51 for 8 *d.f.* Thus, we have established that the covariance structure is not homogeneous in the two treatment groups, a finding which confirms the heterogeneity of covariance structure reported by Zimmerman and Núñez-Antón (1997), who used a classical likelihood ratio test when analysing these data. To investigate the source of this heterogeneity further, we test the omission of the covariate from the correlation structure using  $-2(\hat{\ell}_{II} - \hat{\ell}_I) \sim asym \chi_4^2 = 2$ , a non significant result. However, omitting the covariate from the innovation variance structure using  $-2(\hat{\ell}_{III} - \hat{\ell}_I)$  yields  $\chi_4^2 = 34$ . Accordingly, we adopt Model II which indicates that the innovation variances are different in the two treatment groups.

Figure 1 provides graphical comparisons for several different model fits. In panels (a) and (b), we display the sample log-innovation variances (dot points) and the fitted curves for Model II (solid curve) and Model IV (dash-dot curve), respectively, derived from the entire cattle dataset. Panel (a) shows how Model IV fails to capture the curvature correctly in treatment group A while panel (b) shows that both models provide a similar summary of the data in treatment group B.

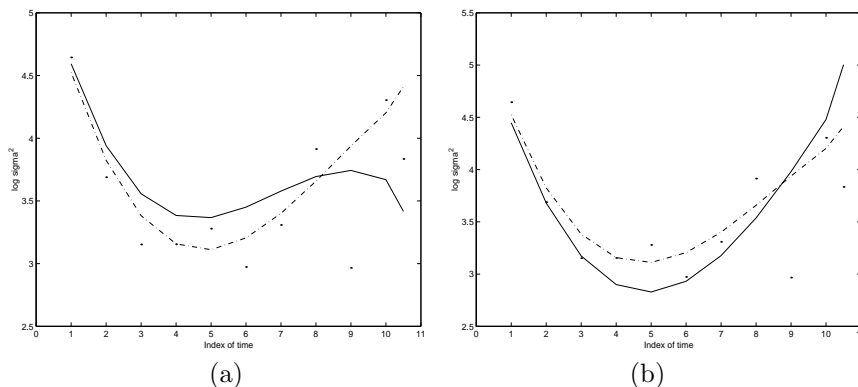


Figure 1. Comparisons for different models: Panels (a) and (b): the sample log-innovation variances for the whole data (dot points) and the fitted curves by Model II (solid curve) and Model IV (dash-dot curve) for Groups A and B respectively.

## 5 Discussion

We have generalised the specification of the marginal covariance matrix employed in classical linear mixed models, from  $\Sigma(t; \theta)$  to  $\Sigma(t, \theta, x, \beta^*)$ . Being *data-driven* the approach is based on observables and *not* on random effects. Usually, the covariance structure in LMMs is assumed not to depend on the covariate space, through fixed effects, but this is simply another model choice. Accordingly, such choices can now be tested in practice.

It may be argued, contrary to the classical assumption (2), that covariates are permissible since in LMM the covariance matrix may be written as  $\Sigma_i = \sigma^2(Z_i B_i Z_i' + W_i)$ . However, covariates which appear in  $\Sigma_i$  are associated with random coefficients and not fixed effects. Moreover, such covariates are constrained to appear as frailty effects in the model for the observations  $Y_i$  in order that their variance components appear in  $\Sigma_i$ , through the conventional marginalisation process. These restrictions do not apply in the more general framework discussed in this paper. For example, the

models for the mean and covariance may depend on different sets of fixed effects. Moreover, the finding, in any analysis, that  $\Sigma_i$  depends on a set of fixed effects automatically disqualifies the conventional Laird-Ware model as a generating mechanism for the observations. Thus, our methods provide a test of the inappropriateness of the Laird-Ware model. We re-emphasise that the entire analysis has been conducted without recourse to random effect assumptions.

In the example analysed, we have demonstrated that the methods developed can detect and explain heterogeneity in the covariance structure and we anticipate that their application will improve insight into the effect of intervention in longitudinal randomised controlled trials.

## References

- Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Appl. Statist.*, **36**, 296-308.
- Laird, N. M. & Ware, J. J. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-74.
- Pan, J. X., Fang, K. T. & von Rosen, D. (1997). Local influence assessment in the growth curve model with unstructured covariance. *J. Statist. Infer. and Plan.*, **62**, 263-278.
- Pan J.X. & MacKenzie G. (2000). Regression models for covariance structure in Longitudinal studies. *Biometrika*, *submitted*
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**, 677-90.
- Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87**, 425-35.
- Zimmerman, D. L. & Núñez-Antón, V. (1997). Structured antedependence models for longitudinal data. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions*, Springer Lecture Notes in Statistics, No. 122, Ed. T. G. Gregoire et al., pp. 63-76. New York: Springer-Verlag.