Modelling Mean-Covariance Structures in the Growth Curve Models

Jianxin Pan and Dietrich von Rosen

Research Report
Centre of Biostochastics

Swedish University of
Agricultural Sciences

Report 2004:4
ISSN 1651-8543

Modelling Mean-Covariance Structures in the Growth Curve Model

Jianxin Pan[1]
Department of Mathematics
University of Manchester

Dietrich von Rosen
Centre of Biostochastics
Swedish University of Agricultural Sciences

Abstract

The growth curve model (GCM) has been widely used in longitudinal studies and repeated measures. Most existing approaches for statistical inference in the GCM assume a specific structure on the within-subject covariances, e.g., compound symmetry, AR(1) and unstructured covariances. This specification, however, may select a suboptimal or even wrong model, which in turn may affect the estimates of regression coefficients and/or bias standard errors of the estimates. Accordingly, statistical inferences of the GCM may be severely affected by misspecification of covariance structures. Within the framework of the GCM in this paper we propose a data-driven approach for modelling the within-subject covariance structures, investigate the effects of misspecification of covariance structures on statistical inferences and study the possible heterogeneity of covariances between different treatment groups.

Keywords: Covariance structures, growth curve models, heterogeneity of covariances, joint mean-covariance modelling, maximum likelihood estimation, misspecification of covariance structures.

AMS 1991 subject classifications: Primary 62F15; Secondary 62F30, 62H12, 62H15

[1]E-mail address to the correspondence author: jpan@maths.man.ac.uk

1 Introduction

The *growth curve models* (GCM) are generalized multivariate analysis-of-variance models that are useful especially in longitudinal studies and repeated measures (Potthoff and Roy, 1964). The GCM is defined by

$$Y_{p\times n} = X_{p\times m} B_{m\times r} Z_{r\times n} + \epsilon_{p\times n} \tag{1}$$

where Y is the response matrix of n subjects measured at p time points, and X and Z are within- and between-subject design matrices with ranks m and r, respectively. Typically, the columns of X are the powers of time at which repeated measures are made when polynomials of time are used to model the mean structures. The rows of Z are the indicators of treatment groups, i.e., the ith row of Z is given by $z_i' = (0, ..., 0, 1'_{n_i}, 0, ..., 0)$ where 1_{n_i} is the $(n_i \times 1)$ vector with all components being one and n_i is the sample size of the ith treatment group ($i = 1, 2, ..., r$; $\sum_{i=1}^{r} n_i = n$). The columns of the error matrix ϵ are assumed to be independent p-variate Normal with $(p \times 1)$ mean vector 0 and $(p \times p)$ covariance matrices Σ_i, depending on the treatment group of which the responses are generated ($i = 1, 2, ..., r$). We denote this by $\epsilon \sim N_{p\times n}(0; \Sigma_1, \Sigma_2, ..., \Sigma_r; I_n)$ where I_n is the identity matrix with size n.

When the covariances are homogeneous, i.e., $\Sigma_1 = \Sigma_2 = ... = \Sigma_r \equiv \Sigma$, the estimates of the parameters B and Σ were discussed in the literature by many authors including Potthoff and Roy (1964), Khatri (1966) and von Rosen (1989) among others. Particularly, if the homogeneous covariance matrix Σ is given then the maximum likelihood estimate (MLE) of B must have an explicit form in terms of *generalized weighted least squares* (GWLS)

$$\widehat{B}(\Sigma) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}YZ'(ZZ')^{-1}. \tag{2}$$

When Σ is unknown, a *two-step* estimation strategy is commonly used to calculate the estimate of B. In other words, we first find an *appropriate* estimate of Σ and then plug it into the GWLS in (2) (e.g., Potthoff and Roy, 1964; Gleser and Olkin, 1970). In particular, when the MLE of Σ is utilized then the resulting GWLS estimate is the MLE of B (Rao, 1965; von Rosen, 1989). Obviously, the GWLS in (2) says that the estimate of B may depend on the estimate of Σ, of which an exception is that Σ has the so-called Rao's *simple covariance structure* (SCS) :

$$\Sigma = X\Gamma X' + Q\Theta Q' \tag{3}$$

where Γ and Θ are $(m \times m)$ and $((p-m) \times (p-m))$ positive definite and Q is orthogonal to X, i.e., $Q'X = 0$ (Rao, 1966). In fact, it can be shown

$\widehat{B}(\Sigma) = \widehat{B}(I_p)$ *if and only if* Σ is of Rao' SCS given in (3) (Kariya, 1985; Pan and Fang, 2002). Typical examples of the SCS include *compound symmetry* and *random regression coefficients* structures (Lee, 1988; Pan and Fang, 2002). Accordingly, the estimate of B is affected by the estimate of Σ unless Σ is within the space of the SCS. On the other hand, since the estimated covariance of $\widehat{B}$ is given by

$$\widehat{\mathrm{Cov}}(\widehat{B}) \equiv \widehat{\mathrm{Cov}}(\mathrm{vec}(\widehat{B})) = c\,[(ZZ')^{-1} \otimes (X'\widehat{\Sigma}^{-1}X)^{-1}] \tag{4}$$

(e.g., von Rosen, 1989) where c is a constant and $\otimes$ denotes the Kronecker product of two matrices, it is obvious that $\widehat{\mathrm{Cov}}(\widehat{B})$ depends on the estimate of Σ even if Σ falls into the space of the SCS. Accordingly, correct estimate of the covariances plays an important role in statistical inference of the GCM.

In the statistical literature the GCM was studied under a variety of assumptions of covariance structures, for example, unstructured covariance (UC) by Potthoff and Roy (1964) and von Rosen (1989), the SCS by Rao (1966) and Lee (1988), compound symmetry structure by Lee (1988), AR(1) by Fujikoshi *et al* (1990) and Lee (1988), random regression coefficients structure by Rao (1966), etc.. With the specification of SCS and UC, statistical diagnostics including outlier and influential observation detections was addressed within likelihood and Bayesian framework by Pan and Fang (2002) and Pan *et al* (1997, 1998, 1999). ¿From inferential and predictive points of view, Lee (1991) and Keramidas and Lee (1995) suggested several selection criteria to choose an appropriate covariance structure from a menu of candidates. This kind of *menu-selection* procedures, however, may not be optimal. For example, when the true covariance structure is not contained in the menu the selected covariance structure, though "best" in some sense, may not be close to the true value. Consequently, statistical inference may be badly affected by the misspecification of covariance structures. On the other hand, the assumption of homogeneous covariances, i.e., $\Sigma_1 = \Sigma_2 = ... = \Sigma_r$, might not be true in practice. For example, in many biological and medical problems the homogeneity assumption does not hold because different treatment groups may have different variations over time. Also, it is not uncommon that within-subject correlation structures may vary from group to group. Accordingly, we hope to establish a mechanism to test whether or not the assumption of homogeneous covariances is true.

In this paper we propose a *data-driven* approach to jointly model the mean and covariance structures for all treatment groups. The approach is based on a modified Cholesky decomposition advocated by Pourahmadi (1999; 2000) for modelling homogeneous covariance structures. We extend Pourahmadi's

(1999) approach to model heterogeneous covariances and in the modelling approach the homogeneity assumption becomes testable. We also investigate the effects of misspecification of covariance structures on statistical inferences in the GCM. This paper is organized as follows. In Section 2 the modified Cholesky decomposition is briefly reviewed and models for mean-covariance structures are proposed. In Section 3 maximum likelihood estimation is developed and in Section 4 principle of testing homogeneous covariances is described. In Section 5 a real data set, Cattle data (Kenward, 1987), is analyzed for illustration. Numerical comparisons between the data-driven and menu-selection approaches are made as well. In Section 6 we discuss some further issue and in the Appendix we give the technical details of the proposed approach.

2 Regression models for mean-covariance structures

For illustration, let us look at the homogeneous GCM $Y \sim N_{p\times n}(XBZ, \Sigma, I_n)$ first. We assume the $(p \times p)$ covariance matrix Σ is positive definite in this paper. Accordingly, there is a unique lower triangular matrix T with 1's as diagonal entries and a unique diagonal matrix D with positive diagonal entries such that $T\Sigma T' = D$. This modified Cholesky decomposition has a transparent statistical interpretation: the below-diagonal entries of T are the negatives of the *autoregressive coefficients*, ϕ_{jk}, in

$$\hat{y}_j = \mu_j + \sum_{k=1}^{j-1} \phi_{jk}(y_k - \mu_k),$$

the linear least squares predictor of y_j based on its predecessors $y_{(j-1)}$, ..., y_1, where $\mu_j = E(y_j)$ and y_j is the jth component of the $(p \times 1)$ response y, the column random variable of Y $(j = 1, 2, ..., p)$. It can be shown that the diagonal entries of D are the *innovation variances* $\sigma_j^2 = \mathrm{Var}(y_j - \hat{y}_j)$ (Pourahmadi, 2000). Obviously, it follows that $\Sigma^{-1} = T'D^{-1}T$.

For the heterogeneous GCM $Y \sim N_{p\times n}(XBZ, \Sigma_1, \Sigma_2, ..., \Sigma_r; I_n)$, we take the modified Cholesky decomposition for each covariance matrix Σ_i, i.e., $T_i\Sigma_i T_i' = D_i$, and then obtain the autoregressive coefficients ϕ_{jki} from the lower triangular matrices T_i and the innovation variances σ_{ji}^2 from the diagonal matrices D_i $(j = 1, 2, ..., p;\ k = 1, 2, ..., j-1;\ i = 1, 2, ..., r)$. In a spirit of Pourahmadi (1999), we propose the following regression models to model the

mean and covariance structures, simultaneously,

$$\mu_{ji} = x_j'\beta_i, \qquad \phi_{jki} = a_{jk}'\gamma_i \qquad \text{and} \qquad \log\sigma_{ji}^2 = h_j'\lambda_i \tag{5}$$

where μ_{ji} is the mean of the responses in the ith group measured at the jth time point, β_i, γ_i and λ_i are $(m\times 1)$, $(q\times 1)$ and $(d\times 1)$ regression coefficients for the ith group, and β_i is actually the ith column of B. The covariates x_j (i.e., the transpose of the jth row of X), a_{jk} and h_j are associated with the powers of time when using polynomials of time to model the mean and covariance structures for growth data. For example, we may choose

$$\begin{array}{rcl} x_j & = & (1, t_j, t_j^2, ..., t_j^{m-1})' \\ a_{jk} & = & (1, (t_j - t_k), (t_j - t_k)^2, ..., (t_j - t_k)^{q-1})' \\ h_j & = & (1, t_j, t_j^2, ..., t_j^{d-1})' \end{array} \tag{6}$$

if the within-subject correlation only depends on the elapsed time, where t_j is the jth time point at which observations are made. In the literature a Brownian motion specified to covariance structures of the GCM was considered by Lundbye-Christensen (1991), which is a special case of the mean-covariance models (5) with the structures (6).

The advantages of the joint regression modelling of mean-covariance structures in (5) are multi-folds, for example, a) it is a data-driven approach that is capable to capture the true structures for mean and covariance, b) the resulted estimates of covariance matrices $\widehat{\Sigma}_i$ are guaranteed to be positive definite, c) the reparameterized regression coefficients have transparent statistical interpretations in terms of autoregressive coefficients and innovation variances (Pourahmadi, 1999, 2000), and d) the assumption of homogeneous covariances becomes testable. We will discuss these issues in more details in the following sections.

3 Maximum likelihood estimation

Denote $Y = (Y_1, Y_2, ..., Y_r)$ and $Z = (Z_1, Z_2, ..., Z_r)$ where Y_i and Z_i are the $(p\times n_i)$ responses and $(r\times n_i)$ between-subject design matrices of the ith group $(i = 1, 2, ..., r)$, respectively. Similar to Pourahmadi (1999), it can be shown that the log-likelihood function $\ell \equiv \ell(B; \gamma_1, ..., \gamma_r; \lambda_1, ..., \lambda_r)$ of the heterogeneous GCM $Y \sim N_{p\times n}(XBZ, \Sigma_1, \Sigma_2, ..., \Sigma_r; I_n)$ modeled with the regression models (5), except a constant being $-(pn/2)\log(2\pi)$, has the following three

representations corresponding to B, $(\gamma_1, ..., \gamma_r)$ and $(\lambda_1, ..., \lambda_r)$, respectively,

$$\begin{aligned}
\ell &= -\tfrac{1}{2}\sum_{i=1}^{r} n_i \log|\Sigma_i| - \frac{1}{2}\sum_{i=1}^{r} \mathrm{tr}\{\Sigma_i^{-1}(Y_i - XBZ_i)(Y_i - XBZ_i)'\} \\
&= -\tfrac{1}{2}\sum_{i=1}^{r} n_i \log|D_i| - \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{p}(e_{ji} - [A^e_{ji}]'\gamma_i)'D_i^{-1}(e_{ji} - [A^e_{ji}]'\gamma_i) \qquad (7) \\
&= -\tfrac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{p} n_i h_j'\lambda_i - \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{p}(e_{ji} - \hat{e}_{ji})'(e_{ji} - \hat{e}_{ji})/\exp\{h_j'\lambda_i\}
\end{aligned}$$

where

$$\begin{aligned}
E_i &= Y_i - XBZ_i = (e_{1i}, e_{2i}, ..., e_{pi})', \\
A^e_{ji} &= \sum_{k=1}^{j-1} a_{jk}e'_{ki} \qquad \text{and} \qquad \hat{e}_{ji} = \sum_{k=1}^{j-1}\phi_{jki}e_{ki}
\end{aligned}$$

for $i = 1, 2, ..., r$ and $j = 1, 2, ..., p$. Note that when $j = 1$ the sum notation $\sum_{k=1}^{0}$ means zero here.

Let $\mathrm{vec}(B) = (\beta_1', \beta_2', ..., \beta_r')'$ be the $mr \times 1$ vector by vectorizing the matrix B through column by column. Taking differentiation of ℓ in (7) with respect to $\mathrm{vec}(B)$, γ_i and λ_i leads to the following estimating equations, respectively,

$$\begin{aligned}
&\partial\ell/\partial\mathrm{vec}(B) = \sum_{i=1}^{r}\mathrm{vec}[X'\Sigma_i^{-1}(Y_i - XBZ_i)Z_i'] = 0 \\
&\partial\ell/\partial\gamma_i = \sum_{j=1}^{p} A^e_{ji}(e_{ji} - [A^e_{ji}]'\gamma_i)/\sigma^2_{ji} = 0 \\
&\partial\ell/\partial\lambda_i = -\tfrac{n_i}{2}\sum_{j=1}^{p} h_j + \frac{1}{2}\sum_{j=1}^{p}(e_{ji} - [A^e_{ji}]'\gamma_i)'(e_{ji} - [A^e_{ji}]'\gamma_i)h_j/\sigma^2_{ji} = 0
\end{aligned} \qquad (8)$$

The estimating equations above in general have no explicit solutions and certain numerical optimization procedures such as the Nowton-Raphson algorithm and Fisher-scoring algorithm are used instead. In the Appendix we show that the Fisher information matrix of the parameter $\theta = (\mathrm{vec}(B)'; \gamma_1', ..., \gamma_r'; \lambda_1', ..., \lambda_r')'$ must have the form

$$\mathcal{I} \equiv E(-\partial^2\ell/\partial\theta\partial\theta') = \begin{pmatrix} \mathcal{I}_{11} & 0 & 0 \\ 0 & \mathcal{I}_{22} & \mathcal{I}_{32}' \\ 0 & \mathcal{I}_{32} & \mathcal{I}_{33} \end{pmatrix} \qquad (9)$$

where

$$\mathcal{I}_{11} \equiv E(-\partial^2\ell/\partial\mathrm{vec}(B)\partial\mathrm{vec}(B)') = \sum_{i=1}^{r}\Big[(Z_iZ_i') \otimes (X'\Sigma_i^{-1}X)\Big] \qquad (10)$$

and the matrices $\mathcal{I}_{22}$ $(rq \times rq)$, $\mathcal{I}_{33}$ $(rd \times rd)$ and $\mathcal{I}_{32}$ $(rd \times rq)$ are block-diagonal with ith block being non-zero $(i = 1, 2, ..., r)$. Their detailed matrix forms are provided in the Appendix.

Based on the above equations, we propose the following Fisher-scoring algorithm to calculate the MLEs of the parameters in the GCM.

Algorithm:

Step 1: Given a starting value of θ, say $\theta_0 = ((\text{vec}(B^0)'; \gamma_1^{0\prime}, ..., \gamma_r^{0\prime}; \lambda_1^{0\prime}, ..., \lambda_r^{0\prime})'$, we form the covariance matrices $\Sigma_i^0 = \Sigma_i(\gamma_i^0, \lambda_i^0)$ using the modified Cholesky decomposition where $i = 1, 2, ..., r$.

Step 2: Use the following procedure

$$\begin{aligned} \text{vec}(B^1) = \text{vec}(B^0) \quad &+ \quad \Big\{ \sum_{i=1}^{r} \Big[(Z_i Z_i') \otimes (X'[\Sigma_i^0]^{-1} X) \Big] \Big\}^{-1} \\ &\times \quad \Big\{ \sum_{i=1}^{r} \text{vec} \Big[X'[\Sigma_i^0]^{-1} (Y_i - X B^0 Z_i) Z_i' \Big] \Big\} \end{aligned}$$

and

$$\begin{pmatrix} \gamma^1 \\ \lambda^1 \end{pmatrix} = \begin{pmatrix} \gamma^0 \\ \lambda^0 \end{pmatrix} + \begin{pmatrix} \mathcal{I}_{22} & \mathcal{I}_{32}' \\ \mathcal{I}_{32} & \mathcal{I}_{33} \end{pmatrix}^{-1}_{\theta=\theta_0} \begin{pmatrix} \partial \ell / \partial \gamma \\ \partial \ell / \partial \lambda \end{pmatrix}_{\theta=\theta_0} \tag{11}$$

to update the parameter estimates of B, $\gamma = (\gamma_1', ..., \gamma_r')'$ and $\lambda = (\lambda_1', ..., \lambda_r')'$, respectively.

Step 3: Use the updated value $\theta_1 = ((\text{vec}(B^1)'; (\gamma^1)'; (\lambda^1)')'$ in Step 2 to replace θ_0 and then repeat Steps 1 and 2 above. These procedures are repeated until convergence for θ.

A by-product of the algorithm above is the asymptotic variance-covariance matrix of the MLE $\hat{\theta} = (\text{vec}(\widehat{B})'; \hat{\gamma}_1', ...\hat{\gamma}_r'; \hat{\lambda}_1', ..., \hat{\lambda}_r')'$, which is obtained by simply calculating the inverse of the Fisher information matrix (9), evaluated at the MLE $\hat{\theta}$. Regarding the starting values γ_i^0 and λ_i^0 $(i = 1, 2, ..., r)$, a convenient choice is $\gamma_1^0 = \cdots = \gamma_r^0 = 0$ and $\lambda_1^0 = \cdots = \lambda_r^0 = 0$ $(i = 1, 2, ..., r)$. In other words, the starting values of covariance matrices in all groups are chosen to be an identity matrix $\Sigma_1 = \cdots = \Sigma_r = I_p$. Alternatively, those can be chosen from the sample covariance matrices (Pourahmadi, 2000). Similarly, the regression coefficients B may start from the sample mean of each group.

4 Hypothesis tests and model selection

As mentioned in Section 1, most literature work in the GCM assumes a homogeneous covariance across all the groups, i.e., $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_r$. Within the framework of the mean-covariance models in (5), this becomes a testable assumption. In fact, testing the homogeneity is equivalent to testing the following hypothesis

$$H_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_r \qquad \text{and} \qquad \lambda_1 = \lambda_2 = \cdots = \lambda_r \tag{12}$$

where the parameters in the regression coefficients B are arbitrary. The likelihood ratio test statistic for testing the homogeneity (12) can be computed straightforwardly but its exact distribution is difficult to obtain. Instead, we could use the asymptotic likelihood ratio test. Let $\hat{\ell}_0$ and $\hat{\ell}_1$ be the maximized log-likelihoods under the null hypothesis H_0 and the alternative hypothesis H_1 of which H_0 is not true, respectively. The homogeneity hypothesis (12) can then be tested using $-2(\hat{\ell}_0 - \hat{\ell}_1) \sim \chi^2$ on $(r-1)(q+d)$ degrees of freedom. When H_0 is rejected, the usual hypothesis of covariance homogeneity is not true and heterogeneous covariances exist across the treatment groups. On the other hand, acceptance of H_0 in (12) implies no evidence to against the homogeneous covariance assumption. Note in this case the homogeneous covariance is modeled jointly with the mean without any specifications of structures. The mean structure, however, may vary from group to group in this case.

When the null hypothesis H_0 in (12) is rejected, we may need to further identify the type of dependence present by investigating the following hypotheses

$$H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_r \qquad \text{and} \qquad H_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_r \tag{13}$$

The first hypothesis in (13) indicates that the innovation variances are the same across the treatment groups, while the second implies there is no difference for within-subject correlation among groups. Again, we can test the null hypotheses in (13) using the asymptotic likelihood ratio tests $-2(\hat{\ell}_0 - \hat{\ell}_1) \sim \chi^2$ on appropriate degrees of freedom.

In the GCM we may also be interested in testing whether or not both the mean and covariance structures are the same across the treatment groups. In other words, we want to test the following hypothesis

$$\begin{aligned} H_0 : \quad & \beta_1 = \beta_2 = \cdots = \beta_r,\ \gamma_1 = \gamma_2 = \cdots = \gamma_r \text{ and} \\ & \lambda_1 = \lambda_2 = \cdots = \lambda_r \end{aligned} \tag{14}$$

where β_i is the ith column of the regression coefficients matrix B ($i = 1, 2, ..., r$). Similarly, the hypothesis in (14) can be tested using the asymptotic

likelihood ratio statistic $-2(\hat{\ell}_0 - \hat{\ell}_1) \sim \chi^2$ on $(r-1)(m+q+d)$ degrees of freedom.

When using polynomials of time to model the mean and covariance structures, obviously we need to choose the appropriate degrees of polynomials m, q and d in (6). In a spirit of Pan and MacKenzie (2003), we propose to use the following Bayesian Information Criterion (BIC)

$$\text{BIC}(m,q,d) = -(2/n)\hat{\ell}_{\max} + (m+q+d)\{(\log n)/n\} \tag{15}$$

to choose the most appropriate degrees of polynomials, where $\hat{\ell}_{\max} = \ell(\hat{B}; \hat{\gamma}_1, ..., \hat{\gamma}_r; \hat{\lambda}_1, ..., \hat{\lambda}_r)$ is the maximized log-likelihood for the models with the specific degree trip (m,q,d) and $m+q+d$ is the number of parameters in the associated models, including polynomials of degree zero (i.e., intercept). The best triple of degrees, say (m^*, q^*, d^*), satisfies

$$(m^*, q^*, d^*) = \arg \min_{(m,q,d)} \{\text{BIC}(m,q,d)\} \tag{16}$$

where m, q and d lie in the range from 1 to p. The global search of the best triple, however, is computationally intensive because the number of maximizations required to find the best triple (m^*, q^*, d^*) is as large as p^3. Even if the number of repeated measurements, p, is mediate, the search for (m^*, q^*, d^*) may be highly computationally time-consuming.

Within the framework of linear regression models, Pan and MacKenzie (2003) proposed a profile search strategy that saturates the degrees m, q and d in pairs. Their study shows that the profile search is able to capture the best triple (m^*, q^*, d^*) in most circumstances. A significant advantage is that the number of maximizations for searching for (m^*, q^*, d^*) reduces to $3p+1$. For more details one can refer to Pan and MacKenzie (2003). In the real data analysis presented in the next section we adopt this strategy to locate the degree triple in the modelling of the mean and covariance structures. Our analysis confirms that the profile search does lead to the global best triple (m^*, q^*, d^*).

So far we have assumed that the degree triple (m,q,d) of polynomials is chosen to be the same across all treatment groups. It is not uncommon, however, that different treatment group may have a different degree triple. In principle the above parameter estimation procedure and model selection strategy are also suitable to this case but the search of the optimal degrees of polynomials is more computational intensive. On the other hand, testing the hypothesis of homogeneous covariance $H_o : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_r$ no longer reduces to testing of the hypothesis (12) in this case because the dimension of

the parameters γ_1, γ_2, ..., γ_r or λ_1, λ_2, ..., λ_r may not be the same. However, the asymptotic likelihood ratio test can be still applied to this case as long as the MLEs of the covariance matrices under the null and alternative hypotheses are obtained.

5 An Example

In this section we analyze Kenward's Cattle data (1987) using the joint mean-covariance modelling strategy. We also compare the data-driven approach to menu-selection methods through the data analysis.

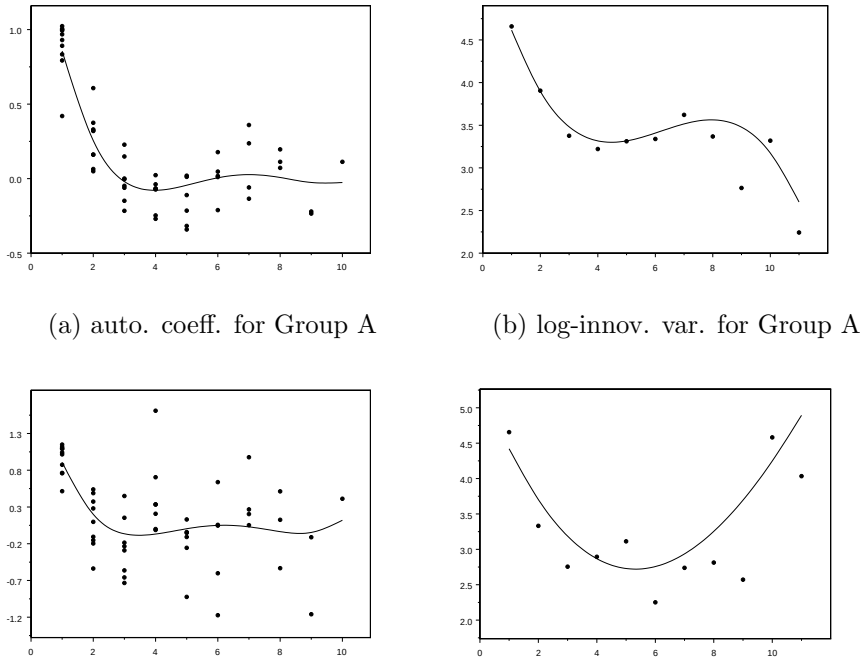
Kenward (1987) analyzed an experiment in which cattle were assigned randomly to two treatment groups A and B, and their weights were recorded to study the effect of treatment on intestinal parasites. Thirty animals received treatment A and another thirty received treatment B. The animals were weighted 11 times over 133-day period at 0, 14, 28, 42, 56, 70, 84, 98, 112, 126 and 133 in days. Pourahmadi (2000) analyzed the data in treatment group A, modelling the covariance structure by adopting a saturated mean model and employing two cubic polynomials of time in the augmented regression model defined in (5), one for the autoregressive coefficients and another for the innovation variances. Below we analyze the two group data simultaneously using the proposed mean-covariance modelling strategy within the framework of growth curve models.

Table 1. *The maximum likelihood estimates of parameters involved in the autoregressive coefficients and innovation variances, i.e.,* $\hat{\gamma}_i' = (\hat{\gamma}_{i1}, ..., \hat{\gamma}_{i5})$ *and* $\hat{\lambda}_i' = (\hat{\lambda}_{i1}, ..., \hat{\lambda}_{i4})$ *(estimated standard errors in parentheses)*

Group	Parameter	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$
B	γ_{1l}	0.185(.006)	-1.628(.104)	1.568(.158)	-1.137(.188)	0.694(.231)
	λ_{1l}	3.518(.077)	0.672(.258)	2.229(.258)	-0.185(.258)	
A	γ_{2l}	0.182(.003)	-1.671(.061)	1.497(.106)	-1.031(.147)	0.365(.164)
	λ_{2l}	3.488(.078)	-1.172(.258)	0.234(.258)	-0.988(.258)	

Firstly, we adopt Pan and MacKenzie's (2003) BIC-based profile search strategy to select the best degree triple (m^*, q^*, d^*) of polynomials used in the modelling. We find that $(m^*, q^*, d^*) = (11, 5, 4)$, i.e., the mean has a saturated structure, and the autoregressive coefficients and innovation variances

are modeled in terms of quadratic and cubic polynomials of lag/time, respectively. The minimum value of the BIC is $\mathrm{BIC}(11, 5, 4) = 72.468$. Table 1 above reports the parameter estimates and the associated standard errors as well, while Figure 1 below gives the sample regressograms (solid points) and the fitted polynomial curves (solid curves) for the autoregressive coefficients and innovation variances in both groups. Note that the estimated coefficients presented in Table 1 are those pertaining to the orthogonal polynomials in order to avoid singularity of the design matrices.



(a) auto. coeff. for Group A

(b) log-innov. var. for Group A

(c) auto. coeff. for Group B

(d) log-innov. var. for Group B

Figure 1. The Sample regressograms (solid points) and the fitted polynomials of lag/time with the best degree triple $(m^*, q^*, d^*) = (11, 5, 4)$. The panels (a) and (b) are those for Group A and (c) and (d) for Group B, respectively.

Secondly, we study whether or not the covariances in Groups A and B are homogeneous, which is equivalent to testing if the null hypothesis (12) is true where $r = 2$. We therefore maximize the log-likelihood functions under the

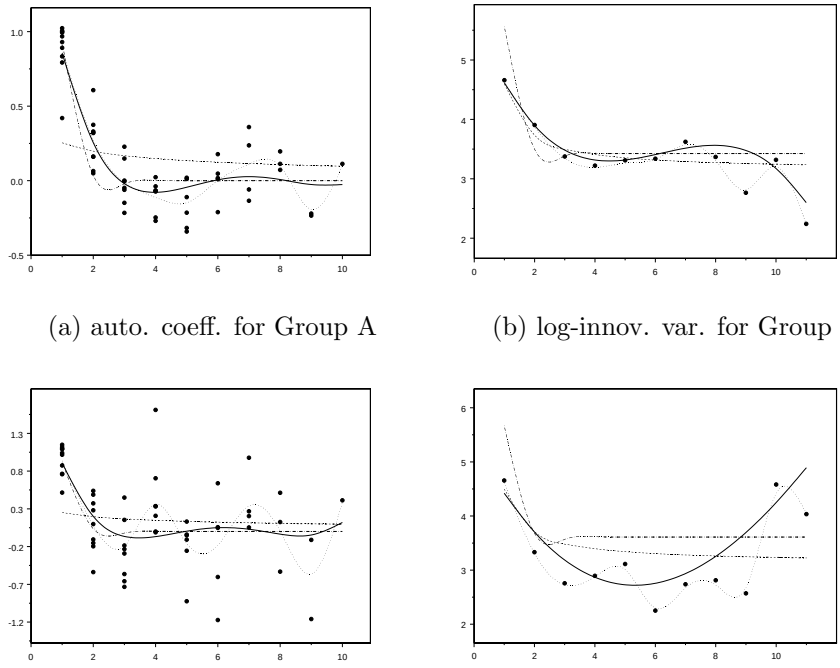
null and alternative hypotheses and obtain $\hat{\ell}_0 = -2120$ and $\hat{\ell}_1 = -2092.167$ so that $-2(\hat{\ell}_0 - \hat{\ell}_1) = 56$. We then compare this value to the Chi-square distribution with $(r-1)(q+d) = 9$ degrees of freedom, i.e., χ^2_9, and conclude that there is a highly significant evidence to against the null hypothesis. In other words, heterogeneous covariances exist for the two group cattle data. Furthermore, we may be concerned with whether either the autoregressive coefficients or innovation variances vary from group to group. We therefore test the hypotheses presented in (13) where $r = 2$. For the innovation variances, we test the null hypothesis $H_0 : \lambda_1 = \lambda_2$ against the alternative hypothesis $H_1 : \lambda_1 \neq \lambda_2$. Under the null and alternative hypotheses, the maximized log-likelihood functions are $\hat{\ell}_0 = -2118$ and $\hat{\ell}_1 = -2092.167$, respectively, and hence the testing statistic value is given by $-2(\hat{\ell}_0 - \hat{\ell}_1) = 58$. When comparing to χ^2_4, we know that the null hypothesis H_0 is rejected and conclude that the innovation variances are different in the two treatment groups. For the autoregressive coefficients, we test the null hypothesis $H_0 : \gamma_1 = \gamma_2$ versus the alternative hypothesis $H_1 : \gamma_1 \neq \gamma_2$. Under H_0 and H_1, the maximized log-likelihood functions are $\hat{\ell}_0 = -2093.113$ and $\hat{\ell}_1 = -2092.167$, respectively, so that the testing statistic takes the value $-2(\hat{\ell}_0 - \hat{\ell}_1) = 4$. When comparing it to χ^2_5, this time we have no evidence to against the null hypothesis H_0. In other words, the autoregressive coefficients are not significantly different in the two treatment groups, which confirms Pan and MacKenzie's (2003) finding where a group indicator is incorporated into covariance modelling.

Table 2. *Comparison between mean-covariance modelling and mean-selections where all models assume a saturated mean*

Covariance	Parameter No.	log-likelihood	BIC
SCS	154	-2018.396	77.789
AR(1)	26	-2161.371	73.520
CS	26	-2409.231	81.782
Modelling	40	-2092.167	72.468

Thirdly, in order to gain an insight of merit on the mean-covariance modelling we compare this strategy to several menu-selection approaches. Table 2 above presents the numbers of parameters, the maximized log-likelihood functions and the BIC values for the mean-covariance modelling and several specifications of covariance structures, including the Rao's simple covariance structure (SCS), AR(1) and and compound symmetry (CS). Note that since the mean structure is saturated, i.e., $m = 11$, the Rao's SCS is completely identical to the unstructured covariance (UC). In Table 2, when AR(1) or CS

is used to specify the covariance structures it is possible that the two treatment groups may have different variances and correlation coefficients. Therefore the BIC values presented in Table 2 are the average in the two groups under these two circumstances. From Table 2 it is obvious that the mean-covariance modelling approach proposed in this paper performs better than the menu-selection approaches in terms of BIC model selection criterion. Figure 2 below compares those modelling approaches through different curves fitted to the sample autoregressive coefficients and innovation variances (dot points) for the two treatment groups, where the solid curve represents the fitting using the mean-covariance modelling technique, while dot, dash and dash-dot curves are the fitting with Rao's SCS, CS and AR(1) covariance specifications, respectively.



(a) auto. coeff. for Group A

(b) log-innov. var. for Group A

(c) auto. coeff. for Group B

(d) log-innov. var. for Group B

Figure 2. The Sample regressograms (solid points), the fitted curves using the mean-covariance modelling technique (solid curve) and with covariance specification being Rao's SCS (dot curve), Compound Symmetry (dash curve) and AR(1) (dash-dot curve). The panels (a) and (b) are those for Group A while (c) and (d) are for Group B.

Again, Figure 2 shows that the mean-covariance modelling approach fits the data well. It also clearly shows the menu-selection approach may misspecify the covariance structures. For example, neither the CS nor AR(1) is able to capture the true covariance structure, while Rao's SCS tends to over-fit the covariance structure. This, in turn, may influence the standard deviation of the estimated regression coefficients and accordingly may bias the statistical inferences of the GCM.

6 Discussion

In this paper we propose a data-driven approach to jointly model the mean and covariance structures for longitudinal data within the framework of growth curve models. The covariance matrices of repeated measures are reparameterized in terms of the modified Cholesky decomposition and the reparameterized parameters have a transparent statistical interpretation - autoregressive coefficients and innovation variances. These reparameterized parameters are further fitted using regression models. The maximum likelihood estimates of the parameters are obtained using the Fisher-scoring algorithm. Based on the joint models, the homogeneous covariance assumption becomes testable. The optimal joint model can be obtained by searching for the most appropriate degree triple of polynomials used for modelling the mean, autoregressive coefficients and innovation variances. A profile BIC-based search strategy is proposed in order to obtain the optimal degree triple.

Compared to menu-selection approaches, the joint mean-covariance modelling strategy specifies no structures on the covariance matrices of within-subject correlation. In contrast, menu-selection approaches assume a specific structure to the covariance matrices. When the structure is misspecified, statistical inferences of the regression coefficients may be incorrect. For example, within the framework of generalized estimating equations (GEE) Wang and Carey (2003) showed that misspecification of covariance structures produces too large standard deviations for regression coefficients and hence results in inefficient estimates. Ye and Pan (2004a) further modeled the mean and covariance structures in GEE using regression models. Very recently they (Ye and Pan, 2004b) proposed to use local-likelihood estimation approach developed by Fan *et al* (1998) to nonparametrically model the mean and covariance structures for large longitudinal data.

For the growth curve models, Rao's simple covariance structure plays a special role in the sense that within this sub-covariance space the MLE of regression coefficients no longer depends on the choice of covariance structures.

It is more interesting to see how this specific covariance structure can be characterized in terms of the autoregressive coefficients and innovation variances. In other words, we want to know under which condition satisfied by the autoregressive coefficients and innovation variances the covariance falls into Rao's simple covariance space. This issue awaits for further exploitation.

To our knowledge, this is the first article that addresses the joint model of mean-covariance structures in the growth curve models in terms of the data-driven regression technique. Of course, under certain specifications of covariance structures statistical modelling was widely discussed in the literature for the GCM, multilevel models, structural equation models, etc. Modelling mean-covariance structures without any specifications of covariance structure distinguishes our approach from the literature work. In addition, our previous experience on statistical diagnostics in the GCM (e.g., Pan and Fang, 2002; Pan *et al*, 1997, 1998, 1999) shows that the covariance structure plays an important role in outlier detection and influential observation identification. The diagnostics issue studied within the framework of joint mean-covariance modelling will be reported in a follow-up paper.

Acknowledgments

We would like to thank Prof Kaitai Fang for his nice collaboration in the area of multivariate analysis, growth curve models and statistical diagnostics during the past decade. Our joint work with him already led to a series of publications including papers (e.g., Pan *et al*, 1997, 1998, 1999) and a Springer Series book (Pan and Fang, 2002). This research work is funded by a EPSRC grant of the UK (GR/R69884).

References

Fan, J., Farmen, M. and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society*, B, **60**, 591-608.

Fujikoshi, Y., Kanda, T. and Tanimura, N. (1990). The growth curve model with an autoregressive covariance structure. *Annals of Institute of Statistical Mathematics* **42**, 533-542.

Gleser, L. J. and Olkin, I. (1970). Linear models in multivariate analysis. In *Essays in Probability and Statistics*. Ed. R.C. Bose, I.M. Chakravarti, P.C. Mahalanobis, C.R. Rao and K.J.C. Smith, pp.267-292, Chapel Hill, North Carolina: University of North Carolina Press.

Kariya, T. (1985). *Testing in the multivariate general linear model*, Kinokuniya Printing Company, Tokyo, Japan.

Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Applied Statistics*, **36**, 296-308.

Keramidas, E. M. and Lee, J. C. (1995). Selection of a covariance structure for growth-curves. *Biometrical Journal*, **37**, 783-797.

Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve. *Annals of Institute of Statistical Mathematics* **18**, 75-86.

Lee, J. C. (1988). Prediction and estimation of growth curve with special covariance structure. *Journal of the American Statistical Association*, **83**, 432-440.

Lee, J. C. (1991). Tests and model selection for the general growth curve model. *Biometrics*, **47**, 147-159.

Lundbye-Christense S. (1991). A multivariate growth curve model for pregnancy. *Biometrics*, **47**, 637-657.

Pan, J. and Fang, K (2002) *Growth curve models and statistical diagnostics*, New York: Springer Series in Statistics.

Pan, J., K. T. Fang and D. von Rosen (1997). Local influence assessment in the growth curve model with unstructured covariance. *Journal of Statistical Planning and Inference*, **62**, 263-278.

Pan, J., K. T. Fang and D. von Rosen (1998). On the posterior distribution of the covariance matrix of the growth curve model. *Statistics & Probability Letters*, **38**, 33-39.

Pan, J., K. T. Fang and D. von Rosen (1999). Bayesian local influence in the growth curve model with unstructured covariance. *Biometrical Journal*, **41**, 641-658.

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

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, **51**, 313-326.

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.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, **52**, 447-458.

Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis. In: P.Krishnaiah, ed., *Multivariate Analysis*, Academic Press, New York, 87-103.

von Rosen, D. (1989). Maximum likelihood estimates in multivariate linear normal model. *Journal of Multivariate Analysis*, **31**, 187-200.

Ye, H. and Pan, J. (2004a). Modelling covariance structures in generalized estimating equations for longitudinal data. *Proceedings of the 19th International Workshop of Statistical Modelling*, Florence, Italy, July, 244-48.

Ye, H. and Pan, J. (2004b). Nonparametrically modelling the mean and covariance structures for large longitudinal data. *Reprint*

Wang, Y. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalized estimating equations performance. *Biometrika* **90**, 29-41.

Appendix: Derivation of the Fisher Information matrix

First, based on (8) we have the second-order derivatives of the log-likelihood with respect to vec(B), γ_i and λ_i as given below

$$\begin{aligned}
\partial^2\ell/\partial\mathrm{vec}(B)\partial\mathrm{vec}'(B) &= -\sum_{i=1}^{r}\Big[(Z_iZ_i')\otimes(X'\Sigma_i^{-1}X)\Big] \\
\partial^2\ell/\partial\gamma_i\partial\gamma_i' &= -\sum_{j=1}^{p}[A_{ji}^e][A_{ji}^e]'/\sigma_{ji}^2 \\
\partial^2\ell/\partial\lambda_i\partial\lambda_i' &= -\tfrac{1}{2}\sum_{j=1}^{p}(e_{ji}-[A_{ji}^e]'\gamma_i)'(e_{ji}-[A_{ji}^e]'\gamma_i)(h_jh_j')/\sigma_{ji}^2
\end{aligned} \tag{A.1}$$

where $A_{ji}^e=\sum_{k=1}^{j-1}a_{jk}e_{ki}'$ and e_{ji}' is the kth row of the residual matrix $E_i=Y_i-XBZ_i$ ($j=1,2,...,p$; $i=1,2,...,r$). Similarly, the second-order mixed derivatives of ℓ with respect to vec(B), γ_i and λ_i can be written into

$$
\begin{aligned}
\partial^2\ell/\partial\gamma_i\partial\text{vec}'(B) &= -\sum_{j=1}^{p}\Big\{[A_{ji}^x][(e_{ji}-[A_{ji}^e]'\gamma_i)'Z_i'\otimes I_m] \\
&\qquad + [A_{ji}^e][Z_i'\otimes(x_j-[A_{ji}^x]'\gamma_i)]\Big\}/\sigma_{ji}^2 \\
\partial^2\ell/\partial\lambda_i\partial\text{vec}'(B) &= -\sum_{j=1}^{p}h_j\text{vec}'\Big[(x_j-[A_{ji}^x]'\gamma_i)(e_{ji}-[A_{ji}^e]'\gamma_i)'Z_i'\sigma_{ji}^2\Big] \\
\partial^2\ell/\partial\lambda_i\partial\gamma_i' &= -\sum_{j=1}^{p}h_j(e_{ji}-[A_{ji}^e]'\gamma_i)'[A_{ji}^e]'/\sigma_{ji}^2
\end{aligned}
\tag{A.2}
$$

where $A_{ji}^x=\sum_{k=1}^{j-1}a_{jk}x_k'$ and $X=(x_1,...,x_p)'$.

Second, when taking expectation to the first equation of (A.1) it obviously results in $\mathcal{I}_{11}$ as given by (10) because the second-order derivative of ℓ with respect to vec(B) is a constant. On the other hand, by noting that $E(e_{ji})=0$ and $E(A_{ji}^e)=\sum_{k=1}^{j-1}a_{jk}E(e_{ki}')=0$ we have $\mathcal{I}_{21}=0$ and $\mathcal{I}_{31}=0$. Therefore, the Fisher information matrix must have the form given in (9), in which the matrices $\mathcal{I}_{22}$, $\mathcal{I}_{32}$ and $\mathcal{I}_{33}$ are obviously block-diagonal due to the forms of score functions given in (8).

The ith diagonal block of $\mathcal{I}_{22}$ $(i=1,2,...,r)$ can be calculated through

$$
\begin{aligned}
\mathcal{I}_{22}(i,i) &= E[-\partial^2\ell/\partial\gamma_i\partial\gamma_i'] = \sum_{j=1}^{p}E(A_{ji}^e[A_{ji}^e]')/\sigma_{ji}^2 \\
&= \sum_{j=1}^{p}E\Big([\sum_{k=1}^{j-1}a_{jk}e_{ki}'][\sum_{l=1}^{j-1}a_{jl}e_{li}']'\Big)/\sigma_{ji}^2 \\
&= \sum_{j=1}^{p}\sum_{k=1}^{j-1}\sum_{l=1}^{j-1}a_{jk}a_{jl}'E(e_{ki}'e_{li})/\sigma_{ji}^2 \\
&= \sum_{j=1}^{p}\sum_{k=1}^{j-1}\sum_{l=1}^{j-1}a_{jk}a_{jl}'(n_i\sigma_{kli})/\sigma_{ji}^2 \\
&= n_i\sum_{j=1}^{p}W_{ji}/\sigma_{ji}^2 = n_iW_i
\end{aligned}
\tag{A.3}
$$

where

$$
W_{ji}=\sum_{k=1}^{j-1}\sum_{l=1}^{j-1}a_{jk}a_{jl}'\sigma_{kli} \qquad \text{and} \qquad W_i=\sum_{j=1}^{p}W_{ji}/\sigma_{ji}^2 \tag{A.4}
$$

where σ_{kli} is the (k,l)th element of the matrix Σ_i. On the other hand, since $T_iE_i=E_i-\widehat{E}_i\sim N_{p\times n_i}(0,D_i,I_{n_i})$ where $\widehat{E}_i=(\hat{e}_{1i},\hat{e}_{2i},...,\hat{e}_{pi})'$ and $\hat{e}_{ji}=$

$\sum_{k=1}^{j-1} \phi_{jki} e_{ki}$ we have $e_{ji} - \hat{e}_{ji} \sim N_{n_i}(0, \sigma_{ji}^2 I_{n_i})$ so that

$$(e_{ji} - \hat{e}_{ji})'(e_{ji} - \hat{e}_{ji})/\sigma_{ji}^2 \sim \chi_{n_i}^2$$

which implies that $E[(e_{ji} - \hat{e}_{ji})'(e_{ji} - \hat{e}_{ji})/\sigma_{ji}^2] = n_i$. Accordingly, the ith diagonal block of $\mathcal{I}_{33}$ $(i = 1, 2, ..., r)$ can be expressed as

$$\begin{aligned}
\mathcal{I}_{33}(i,i) &= E[-\partial^2 \ell / \partial \lambda_i \partial \lambda_i'] \\
&= \tfrac{1}{2} \sum_{j=1}^{p} E\Big[(e_{ji} - [A_{ji}^e]'\gamma_i)'(e_{ji} - [A_{ji}^e]'\gamma_i)/\sigma_{ji}^2\Big](h_j h_j') \\
&= \tfrac{1}{2} \sum_{j=1}^{p} E\Big[(e_{ji} - \hat{e}_{ji})'(e_{ji} - \hat{e}_{ji})/\sigma_{ji}^2\Big](h_j h_j') \\
&= \tfrac{n_i}{2} \sum_{j=1}^{p} (h_j h_j') = \frac{n_i}{2} H'H
\end{aligned} \tag{A.5}$$

where $H = (h_1, h_2, ..., h_p)'$ is the design matrix involved in the modelling of innovation variances.

Finally, the ith diagonal block of $\mathcal{I}_{32}$ is given by

$$\begin{aligned}
\mathcal{I}_{32}(i,i) &= E[-\partial^2 \ell / \partial \lambda_i \partial \gamma_i'] \\
&= \sum_{j=1}^{p} h_j E\Big\{(e_{ji} - [A_{ji}^e]'\gamma_i)'[A_{ji}^e]'\Big\}/\sigma_{ji}^2 \\
&= \sum_{j=1}^{p} h_j \Big(E[A_{ji}^e e_{ji}] - E([A_{ji}^e][A_{ji}^e]')\gamma_i\Big)'/\sigma_{ji}^2 \\
&= n_i \sum_{j=1}^{p} h_j (A_{ji}^\sigma - W_{ji}\gamma_i)'/\sigma_{ji}^2
\end{aligned} \tag{A.6}$$

where $E([A_{ji}^e][A_{ji}^e]') = n_i W_{ji}$ is already showed in (A.3) while

$$\begin{aligned}
E[A_{ji}^e e_{ji}] &= E\Big(\sum_{k=1}^{j-1} a_{jk} e_{ki}' e_{ji}\Big) = \sum_{k=1}^{j-1} a_{ji} E(e_{ki}' e_{ji}) \\
&= \sum_{k=1}^{j-1} a_{ji}(n_i \sigma_{jki}) = n_i \sum_{k=1}^{j-1} a_{jk} \sigma_{jki} \equiv n_i A_{ji}^\sigma.
\end{aligned}$$

In summary, the Fisher information matrix has the form in (9) in which the block forms of $\mathcal{I}_{11}$, $\mathcal{I}_{22}$, $\mathcal{I}_{33}$ and $\mathcal{I}_{32}$ are provided by (10), (A.3), (A.5) and (A.5), respectively.