Estimation and Prediction of Generalized Growth Curve with Grouping Variances in AR(*q*) Dependence Structure

Jack C. Lee[*] and Ying-Lin Hsu

Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

Summary

In this paper we consider maximum likelihood analysis of generalized growth curve model with the Box-Cox transformation when the covariance matrix has AR(*q*) dependence structure with grouping variances. The covariance matrix under consideration is $\Sigma = \mathbf{D}_\sigma \mathbf{C} \mathbf{D}_\sigma$ where $\mathbf{C}$ is the correlation matrix with stationary autoregression process of order q, $q < p$ and $\mathbf{D}_\sigma$ is a diagonal matrix with p elements divided into $g(\leq p)$ groups, i.e., $\mathbf{D}_\sigma$ is a function of $\{\sigma_1, \ldots, \sigma_g\}$ and $-1 < \rho < 1$ and σ_l, $l = 1, \ldots, g$, are unknown. We consider both parameter estimation and prediction of future values. Results are illustrated with real and simulated data.

Key words and phrases: Box-Cox transformation; Heteroscedasticity; Maximum likelihood estimates; Predictions; Simulations

1. Introduction

We consider a generalized multivariate analysis of variance model useful especially for many types of growth curve problems. The model was first proposed by Potthoff and Roy (1964) and subsequently considered by many authors, including Khatri (1966), Geisser (1970, 1980), Lee and Geisser (1972, 1975), Fearn (1975), Jennrich and Schluchter (1986), Rao (1987), Lee (1988, 1991), von Rosen (1991), Pan, Fang and von Rosen (1999), among others.

The growth curve model is defined as

$$\underset{p \times N}{\mathbf{Y}} = \underset{p \times m}{\mathbf{X}} \underset{m \times r}{\boldsymbol{\tau}} \underset{r \times N}{\mathbf{A}} + \underset{p \times N}{\boldsymbol{\epsilon}} \,, \tag{1.1}$$

where $\boldsymbol{\tau}$ is an unknown matrix and $\mathbf{X}$ and $\mathbf{A}$ are known design matrices of ranks $m < p$ and $r < N$, respectively. The columns of $\boldsymbol{\epsilon}$ are independent p-variate normal, with mean vector $\mathbf{0}$ and common covariance matrix Σ. In general, p is the number of time (or spatial) points observed on each of the N individuals, m and r, which specify the degree of polynomial in time (or space) and the number of distinct groups, respectively, are assumed known. The design matrices $\mathbf{X}$ and $\mathbf{A}$ will therefore characterize the degree of the growth function and the distinct

---

* Corresponding author: jclee@stat.nctu.edu.tw

grouping out of the N independent vector observations. Potthoff and Roy (1964) gave many examples of growth curve applications for the model (1.1). Lee and Geisser (1975), Rao (1987) and Lee (1988), among others, applied the model to some biological data. Keramidas and Lee (1990) applied the model to forecast of technology substitutions.

Lee (1988, 1991), Lee and Lu (1987) and Keramidas and Lee (1990, 1995) demonstrated repeatedly the importance of the AR(1) dependence, or serial covariance structure, for the covariance matrix $\mathbf{\Sigma}$ for the model (1.1).

Mansour, Nordheim and Rutledge (1985) allowed for time heteroscedasticity by applying a nonstationary AR(1) error process for repeated measures experiments. Geary (1989) presented a more general case. Rochon (1992) presented stationary ARMA covariance structures with time heteroscedasticity. Laird and Ware (1982) considered the random effects model for longitudinal data, which included growth curve as a special case. They provided for polynomial trends over time, e.g. random slope. Jennrich and Schluchter (1986) provided the information about the use of covariance structures which included autoregressive as well as random effects error terms. In our paper, we provided for grouping the variances in order to construct more parsimonious and realistic covariance structure and provided the Box-Cox transformation in order to make growth function linear or piecewise linear, which is helpful for prediction. We conducted the likelihood ratio test for the adequacy of grouping the variances. In order to conduct the test, we have been aided by the plot of confidence intervals for standard deviations in the model. In theory, however, some of the results given in Laird and Ware (1982) and Jennrich and Schluchter (1986) could be extended to provide the features considered in this paper.

The estimation of parameters and prediction of future values for the model (1.1) having the covariance structure $\mathbf{\Sigma} = \sigma^2 \mathbf{C}$, $\mathbf{C}$ is the correlation matrix with stationary autoregression process of order 1, or AR(1), have been studied using the ML method by Lee (1988). The purpose of this paper is to consider an extension of this covariance structure to an AR(q) process with grouping variances. An AR(q) dependence structure with grouping variances in the growth curve model will be much more general and quite useful in practice because it includes the AR(1) dependence with a common variance as a special case. When the AR(q) dependence structure with grouping variances holds for $\mathbf{\Sigma}$, we have

$$\mathbf{\Sigma} = \mathbf{D}_\sigma \mathbf{C} \mathbf{D}_\sigma \,, \tag{1.2}$$

where $\mathbf{C} = (c_{ij})$ for $i, j = 1, \ldots, p$, i.e.,

$$c_{ij} = \begin{cases} \rho_{|i-j|} \,, & \text{when } i \neq j \,, \\ 1 \,, & \text{when } i = j \,, \end{cases} \tag{1.3}$$

$$\rho_i = \sum_{j=1}^{q} \phi_j \rho_{|i-j|} \,, \quad \text{for } i \geq 1 \,, \tag{1.4}$$

$\mathbf{D}_\sigma$ is a $p \times p$ diagonal matrix and function of $\{\sigma_1, \ldots, \sigma_g\}$ and $\phi_j, j = 1, \ldots, q$ are unknown and satisfy the assumption that the roots of the equation $1 - \sum\limits_{j=1}^{q} \phi_j x^j = 0$ lie outside the unit circle $\|x\| = 1$ in the complex plane and $\sigma_l > 0$ for $l = 1, \ldots, g$ are unknown. In the grouping of the variances, we assume that there are g_i members in the ith group for $i = 1, \ldots, g$. The purpose of this paper is to consider this covariance structure from the maximum likelihood point of view hoping that a more effective and practical solution can be furnished. We will compare our results with those based on the ML method for the AR(1) dependence as given in LEE (1988) via real and simulated data. Since $\sigma_i > 0$, there is a one-to-one correspondence between σ_i and σ_i^2. Thus, for convenience, we will be dealing with σ_i instead of σ_i^2 mathematically for the rest of this paper. Also, it is trivial to note that equality of variances is equivalent to equality of standard deviations.

In addition to the inferences on the parameters $\boldsymbol{\tau}$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)'$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_g)'$, we will also consider conditional prediction problem for the growth curve model as specified by (1.1)–(1.4). Let $\mathbf{V}$ be $p \times K$ future observations drawn from the generalized growth curve model; that is, the set of future observations is such that given the parameters $\boldsymbol{\tau}$ and $\boldsymbol{\Sigma}$,

$$E(\mathbf{V}) = \mathbf{X\tau F}, \tag{1.5}$$

where E( ) denotes expected value, $\mathbf{F}$ is a known $r \times K$ matrix, and the columns of $\mathbf{V}$ are independent and multivariate normal with a common covariance matrix $\boldsymbol{\Sigma}$. It is noted that $\mathbf{F} = 1$ in the special situation in which $r = K = 1$. LEE and GEISSER (1972, 1975), FEARN (1975), and LEE (1988, 1991) considered the problem of predicting $\mathbf{V}^{(2)}$, given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$, if $\mathbf{V}$ is partitioned as $\mathbf{V} = (\mathbf{V}^{(1)'}, \mathbf{V}^{(2)'})'$, where $\mathbf{V}^{(i)}$ is $p_i \times K$ $(i = 1, 2)$ and $p_1 + p_2 = p$. If p is interpreted as the number of points in time being observed, then the problem is mainly concerned with predicting the generalized growth curve for future values for the same set of p time points, or a subset of size p_2. When $p_2 < p$ and $K = 1$, it is also called the conditional prediction of the unobserved portion of a partially observed vector.

Meanwhile, we will also consider the situation in which the Box-Cox transformations are applied to the observations $\mathbf{Y}$ in (1.1). The model, called data-based transformed (DBT) model, is

$$\mathbf{Y}^{(\lambda)} = \mathbf{X\tau A} + \boldsymbol{\epsilon}, \tag{1.6}$$

where $\mathbf{Y}^{(\lambda)} = (\mathbf{Y}_{ij}^{(\lambda)})$ and

$$\mathbf{Y}_{ij}^{(\lambda)} = \begin{cases} \{(\mathbf{Y}_{ij} + \gamma)^\lambda - 1\}/\lambda, & \text{when } \lambda \neq 0 \\ \ln(\mathbf{Y}_{ij} + \gamma), & \text{when } \lambda = 0 \end{cases}, \tag{1.7}$$

γ is assumed to be a known constant such that $\mathbf{Y}_{ij} + \gamma > 0$ for all i and j, λ is an unknown parameter, and columns of $\boldsymbol{\epsilon}$ are independently and normally distributed with mean vector $\mathbf{0}$ and common covariance matrix $\boldsymbol{\Sigma}$ as specified in (1.2)–(1.4). It is noted that λ could assume different values for different segments of observations. In

LEE and LU (1987), tremendous improvement was found in predictive accuracy using the DBT model for technology substitutions. This is primarily due to the fact that the linearity assumption for the growth function can be enhanced significantly with the Box-Cox transformation, along with incorporating into the model the proper dependence structure among the observations. KERAMIDAS and LEE (1990) combined the concepts of the Box-Cox transformation and a generalized growth curve model with serial structure for forecasting technological substitutions based on the maximum likelihood method when repeated measurements of short time series are available. If the growth function is not linear, we can use the Box-Cox transformation so that the function is linear or piecewise linear, but the transformation may not stabilize the variance. Thus, it is appropriate to construct grouping of variances with AR(q) dependence.

In Section 2, maximum likelihood estimation of parameters and prediction of future values are considered. In Section 3, we derive asymptotic result and test for the equality and grouping of variances. The results are illustrated in Section 4 with two real data sets and a simulation. Some concluding remarks are given in Section 5. Finally, the derivation of the Hessian information matrix is given in the Appendix.

2. Parameter estimation and prediction of $\mathbf{V}^{(2)}$ based on the maximum likelihood theory

From equation (1.4), we obtain the following equations:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_q\rho_{k-q}, \quad \text{for } k > 0, \tag{2.1}$$

where for $j < 0$, $\rho_j = \rho_{-j}$ and $\rho_0 = 1$. From equation (2.1), for $k = 1, \ldots, q$, we obtain the well-known Yule-Walker equations by replacing ρ_j with $\hat{\rho}_j$. The Yule-Walker estimate solves these equations for $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)'$ as a function of ρ's. Here, for the constraints on $\boldsymbol{\rho}$, we would express $\boldsymbol{\rho}$ as a function of $\boldsymbol{\phi}$, the parameters to be estimated. From equation (2.1), if q is odd,

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_{\frac{q+1}{2}-1}\rho_1 + \phi_{\frac{q+1}{2}}\rho_0 + \phi_{\frac{q+1}{2}+1}\rho_{-1}$$
$$+ \ldots + \phi_q\rho_{k-q};$$

else,

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_{\frac{q}{2}-1}\rho_1 + \phi_{\frac{q}{2}}\rho_0 + \phi_{\frac{q}{2}+1}\rho_{-1} + \ldots + \phi_q\rho_{k-q}.$$

Since $\rho_j = \rho_{-j}$ and $\rho_0 = 1$, it follows that

$$\rho_k = \begin{cases} \phi_k + \sum_{j=1}^{k-1}(\phi_j + \phi_{2k-j})\,\rho_{k-j}, & \text{if } k = \dfrac{q+1}{2} \\[2ex] \phi_k + \sum_{j=1}^{k-1}(\phi_j + \phi_{2k-j})\,\rho_{k-j} + \sum_{j'=2k}^{q}\phi_{j'}\rho_{j'-k}, & \text{if } k < \dfrac{q+1}{2} \\[2ex] \phi_k + \sum_{j=1}^{2k-q-1}\phi_j\rho_{k-j} + \sum_{j'=k+1}^{q}(\phi_{2k-j'} + \phi_{j'})\,\rho_{j'-k}, & \text{if } k > \dfrac{q+1}{2}. \end{cases} \tag{2.2}$$

Hence, we have natural constraints on $\boldsymbol{\rho}$ that can be expressed as

$$\mathbf{G}\boldsymbol{\rho} + \boldsymbol{\phi} = \mathbf{0} \tag{2.3}$$

where the matrix $\mathbf{G}$ is a function of $\boldsymbol{\phi}$, element-wise

$$\mathbf{G}_{ij} = \phi_{i+j} + \phi_{i-j} - \delta_{ij}, \tag{2.4}$$

with $\phi_k \equiv 0$ if $k \notin \{1, \ldots, q\}$ and δ_{ij} equals to 1 if $i = j$ and 0 otherwise. Hence, for the AR(q) dependence structure, $\mathbf{C}$ in (1.3), could be expressed as a function of $\boldsymbol{\phi}$.

From model (1.6), we have the following likelihood of $\boldsymbol{\tau}$, $\boldsymbol{\sigma}$, $\boldsymbol{\phi}$ and λ:

$$L(\boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\phi}, \lambda \mid \mathbf{Y}) \propto |\boldsymbol{\Sigma}|^{-\frac{N}{2}} \exp\left\{ -\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} (\mathbf{Y}^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}\mathbf{A})\, (\mathbf{Y}^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}\mathbf{A})' \right\} J_{\mathbf{Y}}, \tag{2.5}$$

where $J_{\mathbf{Y}}$, the Jacobian of the transformation from $\mathbf{Y}^{(\lambda)}$ to $\mathbf{Y}$, is defined as

$$J_{\mathbf{Y}} = \prod_{j=1}^{N} \prod_{i=1}^{p} (\mathbf{Y}_{ij} + \gamma)^{\lambda - 1}. \tag{2.6}$$

From (3.6) of Lee (1988), we have

$$\begin{aligned} &\operatorname{tr} \boldsymbol{\Sigma}^{-1}(\mathbf{Y}^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}\mathbf{A})\, (\mathbf{Y}^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}\mathbf{A})' \\ &= \operatorname{tr} (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}) \left[(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} + (\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})\, \mathbf{A}\mathbf{A}'(\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})' \right] \\ &\quad + \operatorname{tr} (\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}^{(\lambda)}\mathbf{Y}^{(\lambda)'}\mathbf{Z} \end{aligned}$$

and using the fact

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1} \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1} + \mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z})^{-1} \mathbf{Z}',$$

where

$$\mathbf{S} = \mathbf{Y}^{(\lambda)}(\mathbf{I} - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A})\, \mathbf{Y}^{(\lambda)'} \tag{2.7}$$

and $\mathbf{Z}$ is a known $p \times (p - m)$ matrix with rank $p - m$ such that $\mathbf{X}'\mathbf{Z} = \mathbf{0}$, and $\hat{\boldsymbol{\tau}}$, the MLE of $\boldsymbol{\tau}$, is

$$\hat{\boldsymbol{\tau}}(\boldsymbol{\Sigma}) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}^{(\lambda)}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}, \tag{2.8}$$

with

$$\boldsymbol{\Sigma} = \mathbf{D}_{\sigma}\mathbf{C}\mathbf{D}_{\sigma}. \tag{2.9}$$

The MLEs of $\boldsymbol{\sigma}$, $\boldsymbol{\phi}$ and λ, denoted by $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_g)'$, $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_q)'$ and $\hat{\lambda}$, respectively, are obtained by maximizing the profile likelihood function

$$\begin{aligned} &L_{\max}(\boldsymbol{\sigma}, \boldsymbol{\phi}, \lambda) \\ &= |\boldsymbol{\Sigma}|^{-N/2} \exp\left\{ -\tfrac{1}{2} \operatorname{tr} \left[\boldsymbol{\Sigma}^{-1}\mathbf{S} + \mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}^{(\lambda)}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1} \mathbf{A}\mathbf{Y}^{(\lambda)'} \right] \right\} J_{\mathbf{Y}}. \end{aligned} \tag{2.10}$$

Comments regarding the λ values can be applied when there are different values of λ for different segments of observations.

We consider the conditional prediction of the future matrix $\mathbf{V}^{(2)}$, given $\mathbf{Y}$ and $\mathbf{V}^{(1)}$. Schematically, the matrices $\mathbf{Y}$, $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$ are shown below:

$$
\begin{array}{|c|c|}
\hline
 & \mathbf{V}^{(1)} \\
\cline{2-2}
\mathbf{Y} & \mathbf{V}^{(2)} \\
\hline
\end{array}
$$

where $\mathbf{Y}$, $p \times N$, is the complete sample; $\mathbf{V}^{(1)}$, $p_1 \times K$, is the partially observed matrix and $\mathbf{V}^{(2)}$, $p_2 \times K$, is the unobserved portion to be predicted. Of course, $p_1 + p_2 = p$.

We restrict attention to the special situation in which $K = 1$ and $r = 1$; that is, only one vector is being predicted in the same group. This is similar to Lee and Geisser (1975). The approximate mean (denoted by $\hat{\mathbf{V}}^{(2)}$), when the covariance structure given by (1.2) holds, can be obtained as

$$
\hat{\mathbf{V}}^{(2)} = \begin{cases} \{\mathbf{1} + \hat{\lambda}[\mathbf{X}^{(2)}\hat{\boldsymbol{\tau}} + \hat{\boldsymbol{\Sigma}}_{21}\hat{\boldsymbol{\Sigma}}_{11}^{-1}(\mathbf{V}^{(1)}(\hat{\lambda}) - \mathbf{X}^{(1)}\hat{\boldsymbol{\tau}})]\}^{1/\hat{\lambda}} - \gamma, & \text{if } \hat{\lambda} \neq 0 \\ \exp[\mathbf{X}^{(2)}\hat{\boldsymbol{\tau}} + \hat{\boldsymbol{\Sigma}}_{21}\hat{\boldsymbol{\Sigma}}_{11}^{-1}(\mathbf{V}^{(1)}(\hat{\lambda}) - \mathbf{X}^{(1)}\hat{\boldsymbol{\tau}})] - \gamma, & \text{if } \hat{\lambda} = 0\,, \end{cases}
\tag{2.11}
$$

and applying the delta method to obtain an estimate of the approximate covariance matrix of $\hat{\mathbf{V}}^{(2)}$ as follows:

$$
\hat{\boldsymbol{\Sigma}}_{\hat{\mathbf{V}}^{(2)}} \cong \begin{cases} \operatorname{diag}\left((\hat{\lambda}\mathbf{X}^{(2)}\hat{\boldsymbol{\tau}} + \mathbf{1})^{1/\hat{\lambda}-1}\right) \cdot \hat{\boldsymbol{\Omega}} \cdot \operatorname{diag}\left((\hat{\lambda}\mathbf{X}^{(2)}\hat{\boldsymbol{\tau}} + \mathbf{1})^{1/\hat{\lambda}-1}\right), & \text{if } \hat{\lambda} \neq 0 \\ \operatorname{diag}\left(\exp(\mathbf{X}^{(2)}\hat{\boldsymbol{\tau}})\right) \cdot \hat{\boldsymbol{\Omega}} \cdot \operatorname{diag}\left(\exp(\mathbf{X}^{(2)}\hat{\boldsymbol{\tau}})\right), & \text{if } \hat{\lambda} = 0\,, \end{cases}
\tag{2.12}
$$

and

$$
\hat{\boldsymbol{\Omega}} = \frac{1}{N}\left(\mathbf{X}^{(2)} - \hat{\boldsymbol{\Sigma}}_{21}\hat{\boldsymbol{\Sigma}}_{11}^{-1}\mathbf{X}^{(1)}\right)\left(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{(2)} - \hat{\boldsymbol{\Sigma}}_{21}\hat{\boldsymbol{\Sigma}}_{11}^{-1}\mathbf{X}^{(1)}\right)' + \hat{\boldsymbol{\Sigma}}_{21}\hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\Sigma}}_{12}\,,
$$

where $\mathbf{V}^{(1)(\lambda)}$ is defined in manners similar to $\mathbf{Y}^{(\lambda)}$ given in (1.6) and (1.7), $\hat{\boldsymbol{\tau}}$ is as given in (2.7) with $\boldsymbol{\Sigma}$ replaced by $\hat{\boldsymbol{\Sigma}}$, $\mathbf{X}$ is partitioned as $\mathbf{X} = (\mathbf{X}^{(1)'}, \mathbf{X}^{(2)'})'$, $\mathbf{X}^{(i)}$ is $p_i \times m (i = 1, 2)$, $\operatorname{diag}(\mathbf{a})$ is a diagonal matrix with elements $\mathbf{a}$ and

$$
\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{D}}_\sigma \hat{\mathbf{C}} \hat{\mathbf{D}}_\sigma = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{11} & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}_{21} & \hat{\boldsymbol{\Sigma}}_{22} \end{pmatrix},
\tag{2.13}
$$

where $\hat{\boldsymbol{\Sigma}}_{ij}$ is of dimension $p_i \times p_j$ $(p_1 + p_2 = p)$. It is noted that $\mathbf{x}^a = (\mathbf{x}_1^a, \ldots, \mathbf{x}_p^a)'$ where $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)'$ and a is a scalar.

3. Asymptotic result and testing equality and grouping of variances

In this section, we will derive the information matrix which is useful for obtaining the standard errors of the MLEs and testing equality and grouping of variances.

3.1 *Information matrix*

For ease of presentation and some practical consideration, we will restrict our attention to the special situation in which $r = 1$. The log-likelihood obtained from (2.5) is:

$$l = l(\boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\phi}, \lambda \mid \mathbf{Y})$$

$$= -\frac{N}{2} \log(|\boldsymbol{\Sigma}|) - \frac{1}{2} \sum_{i=1}^{N} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}) + \log(J_{\mathbf{Y}}). \quad (3.1)$$

We are unable to take the expectation of the Hessian matrix, which is too complicated. Only observed Hessian matrix is used in the paper. This is still useful in computing the standard errors of the MLEs. Let $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\sigma}', \boldsymbol{\phi}', \lambda)'$. In case there are g groups with g_i members in the group. The Hessian matrix, $\mathbf{H} = \partial^2 l / \partial \boldsymbol{\theta} \, \partial \boldsymbol{\theta}'$, has the following block partitioned form

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{H}_{13} & \mathbf{H}_{14} \\ \mathbf{H}_{12}' & \mathbf{H}_{22} & \mathbf{H}_{23} & \mathbf{H}_{24} \\ \mathbf{H}_{13}' & \mathbf{H}_{23}' & \mathbf{H}_{33} & \mathbf{H}_{34} \\ \mathbf{H}_{14}' & \mathbf{H}_{24}' & \mathbf{H}_{34}' & \mathbf{H}_{44} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial^2 l}{\partial \boldsymbol{\tau} \, \partial \boldsymbol{\tau}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\tau} \, \partial \boldsymbol{\sigma}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\tau} \, \partial \boldsymbol{\phi}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\tau} \, \partial \lambda} \\ \dfrac{\partial^2 l}{\partial \boldsymbol{\sigma} \, \partial \boldsymbol{\tau}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\sigma} \, \partial \boldsymbol{\sigma}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\sigma} \, \partial \boldsymbol{\phi}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\sigma} \, \partial \lambda} \\ \dfrac{\partial^2 l}{\partial \boldsymbol{\phi} \, \partial \boldsymbol{\tau}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\phi} \, \partial \boldsymbol{\sigma}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\phi} \, \partial \boldsymbol{\phi}'} & \dfrac{\partial^2 l}{\partial \boldsymbol{\phi} \, \partial \lambda} \\ \dfrac{\partial^2 l}{\partial \lambda \, \partial \boldsymbol{\tau}'} & \dfrac{\partial^2 l}{\partial \lambda \, \partial \boldsymbol{\sigma}'} & \dfrac{\partial^2 l}{\partial \lambda \, \partial \boldsymbol{\phi}'} & \dfrac{\partial^2 l}{\partial \lambda \, \partial \lambda} \end{pmatrix},$$

$$(3.2)$$

where $\mathbf{H}_{11} = -N\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$ and other derivations are placed in the Appendix.

Thus, an approximate information matrix is

$$\mathbf{I}_{\boldsymbol{\theta}} \cong -\mathbf{H}(\hat{\boldsymbol{\theta}}), \quad (3.3)$$

where $\hat{\boldsymbol{\theta}}$ is the mode of the likelihood function. Under some regularity conditions the estimator $\hat{\boldsymbol{\theta}}$ is asymptotically normal with mean $\boldsymbol{\theta}$ and covariance matrix $\mathbf{I}_{\boldsymbol{\theta}}^{-1}$.

3.2 *Testing equality and grouping of variances*

We will first test the equality of g different variances under the assumption of grouping variances, that is,

$$H_0 : \mathbf{D}\boldsymbol{\sigma} = \mathbf{0} \text{ versus } H_1 : \mathbf{D}\boldsymbol{\sigma} \neq \mathbf{0}$$

where

$$\underset{(g-1) \times g}{\mathbf{D}} = \begin{pmatrix} 1 & -1 & 0 & \dots & & 0 \\ 0 & 1 & -1 & 0 & & \vdots \\ \vdots & 0 & \ddots & \ddots & & 0 \\ 0 & \dots & 0 & & 1 & -1 \end{pmatrix}.$$

The procedure is based on the likelihood ratio criterion,

$$\lambda(\mathbf{Y}) = \frac{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta} \mid \mathbf{Y})}{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta} \mid \mathbf{Y})},$$

and λ_α is chosen so that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} Pr\{\lambda(\mathbf{Y}) \leq \lambda_\alpha; \boldsymbol{\theta}\} \leq \alpha,$$

where $\boldsymbol{\Theta}_0$ is the null space when H_0 is true while $\boldsymbol{\Theta}$ is the entire parameter space under the assumption of grouping variances. Let $Q(\mathbf{Y}) = -2 \ln \lambda(\mathbf{Y})$. Under some mild regularity conditions, the statistic $Q(\mathbf{Y})$ is asymptotically Chi-square distributed with $g - 1$ degrees of freedom under the null hypothesis.

The likelihood ratio test procedure can also be applied to test the null hypothesis

H_0^*: grouping variances versus H_1^*: all p variances are distinct.

The likelihood ratio criterion is

$$\lambda^*(\mathbf{Y}) = \frac{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^*} L(\boldsymbol{\theta} \mid \mathbf{Y})}{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}^*} L(\boldsymbol{\theta} \mid \mathbf{Y})},$$

where $\boldsymbol{\Theta}_0^*$ is the null space when H_0^* is true while $\boldsymbol{\Theta}^*$ is the entire parameter space when the p variances are distinct. Let $Q^*(\mathbf{Y}) = -2 \ln \lambda^*(\mathbf{Y})$. Then under some mild regularity conditions, $Q^*(\mathbf{Y})$ is asymptotically Chi-square distributed with $p - g$ degrees of freedom under the null hypothesis.

4. Illustrative example

This section is devoted to the illustration of the conditional prediction of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$. For the conditional prediction, we will set $K = 1$ and $p_2 = 1, 2, 3, 4$, that is, we will predict the last four observations of a partially observed vector. If $p = 7$ and $p_2 = 1, 2, 3, 4$, we denote the predictions as V_7, $V_6 - V_7$, $V_5 - V_7$ and $V_4 - V_7$, respectively. For the parameter $\boldsymbol{\tau}$, we will assume that the N observations have come from the same group. We will use the predictive sample reuse (PSR), or the leave-one-out (LOO) method. The method was used by Lee and Geisser (1975), Rao (1987) and Lee (1988), among others. The discrepancy measure is the mean absolute relative derivation (MARD) of the predictions from the actuals. The discrepancy measure is defined as

$$\text{MARD} = (Np_2)^{-1} \sum_{j=1}^{p_2} \sum_{i=1}^{N} \frac{|\mathbf{Y}_{ij}^{(2)} - \hat{\mathbf{Y}}_{ij}^{(2)}|}{\mathbf{Y}_{ij}^{(2)}},$$

where $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$, $\mathbf{Y}_i = \begin{pmatrix} \mathbf{Y}_i^{(1)} \\ \mathbf{Y}_i^{(2)} \end{pmatrix}$, $\mathbf{Y}_i^{(1)}$ is $p_1 \times 1$, $\mathbf{Y}_i^{(2)}$ is $p_2 \times 1$, $p_1 + p_2 = p$ and $\hat{\mathbf{Y}}_i^{(2)}$ is its predicted values. It is noted that when $\mathbf{Y}_i^{(2)}$ is being predicted, the complete sample is $\mathbf{Y}_{(i)} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_{i-1}, \mathbf{Y}_{i+1}, \ldots, \mathbf{Y}_N)$ and the partially observed vector is $\mathbf{Y}_i^{(1)}$.

4.1 *The mice data*

The mice data set was reported by WILLIAMS and IZENMAN (1981) and analyzed by RAO (1984, 1987) and LEE (1988, 1991). It consists of weights of 13 male mice, measured at intervals of 3 days, over the 21 days from birth to weaning. The data are plotted in Figure 1. We assume the second degree polynomial for the growth curve, as noted by RAO (1984). When the growth function is quadratic, the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 4 & 9 & 16 & 25 & 36 & 49 \end{pmatrix}',$$

and the design matrix $\mathbf{A}$ is 1×13 vector consisting of all 1s.

The ML estimates with standard errors in parentheses for the model (1.6), with $q = 1$, $m = 3$ and distinct variances, are given in Table 1. When the same λ is applied to all observations for each mouse, the estimate of $\boldsymbol{\sigma}$ with 95% confidence interval for each time point is plotted in Figure 2. We found that the mice data
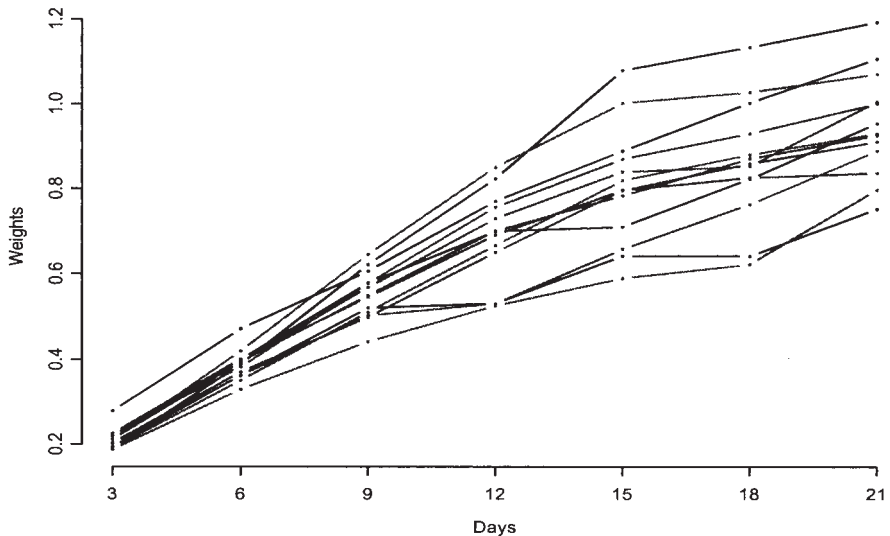

Fig. 1. Plot of mice data

Table 1

ML estimation results for the distinct variances with AR(1) errors

$\hat{\tau}$	−1.1125	0.2507	−0.0152				
	(0.0090)	(0.0085)	(0.0011)				
$\hat{\sigma}$	0.0333	0.0461	0.0666	0.1093	0.1279	0.1333	0.1100
	(0.0099)	(0.0113)	(0.0144)	(0.0226)	(0.0242)	(0.0266)	(0.0198)
$\hat{\rho}$	0.9006						
	(0.054)						
$\hat{\lambda}$	0.8372						
	(0.1420)						

with the same Box-Cox transformation for different time points have a possible change point at day 15. From Table 1, we also found that standard deviation varies with time and stabilize at 12 days, which is also displayed in Figure 2. Hence, a possible model is to group the last three variances as a single group and the rest as four different variances. To test for the adequacy of this grouping variances structure, the value of the log-likelihood function is 169.19 under AR(1) dependence with the above grouping variances. Under the equality of variances, the value of the log-likelihood function is 126.52, resulting in a likelihood ratio statistic $Q = 85.34$ and p-value < 0.0001, indicating that the homogeneity of variances is not acceptable. Meanwhile, we can test the grouping variances (H_0^*) against the alternative hypothesis that all p variances are distinct (H_1^*). When all 7 variances are distinct, the value of the log-likelihood function is 170.10, resulting
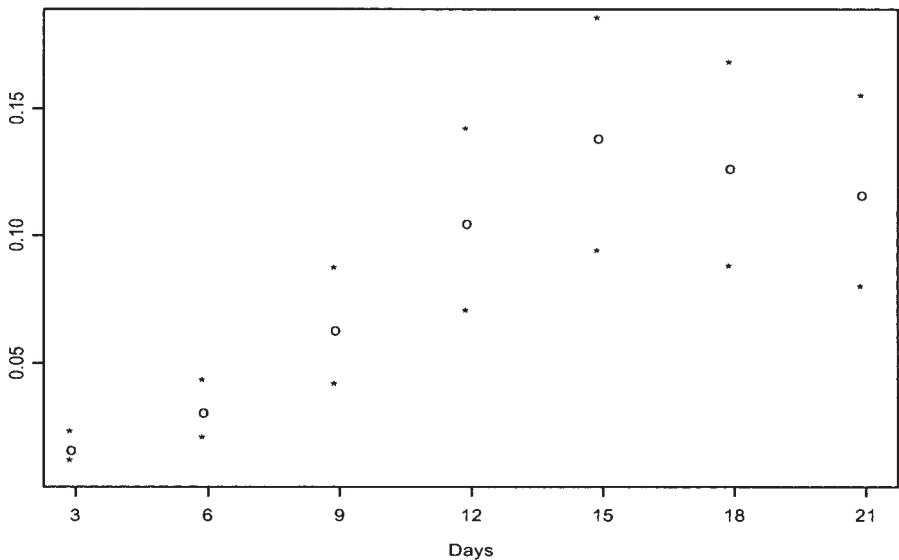

Fig. 2. The estimate of σ of mice data with 95% confidence interval for each time point

Table 2

The MARD between the predicted and the actual values of $\mathbf{V}^{(2)}$: mice data

	Covariance structure	V_7	$V_6 - V_7$	$V_5 - V_7$	$V_4 - V_7$
linear	$\{(1)(2)(3)(4)(5,6,7)\}$ & AR(1)	0.0477	0.0602	0.0617	0.0808
linear	$\{(1,2,3,4,5,6,7)\}$ & AR(1)	0.0454	0.0822	0.0866	0.1097
quadratic	$\{(1)(2)(3)(4)(5,6,7)\}$ & AR(1)	0.0403	0.0523	0.0588	0.0722
quadratic	$\{(1,2,3,4,5,6,7)\}$ & AR(1)	0.0406	0.0525	0.0628	0.0965

in a likelihood ratio statistic $Q^* = 1.82$ and p-value $= 0.5975$, indicating that the grouping of variances is acceptable. Thus, the AR(1) dependence with the above grouping of variances is a possible model for the data.

The comparisons of predictive performance of different grouping variances structures for conditional predictions of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$ without the Box-Cox transformation and with the Box-Cox transformation are summarized in Table 2 and Table 3, respectively. In the tables, the notation $\{(1)\,(2)\,(3)\,(4)\,(5,6,7)\}$ denotes the grouping of the variances corresponding to the seven time points of mice data into five groups with $\sigma_5 = \sigma_6 = \sigma_7$. We found that the grouping variances are better than the common variance with AR(q) dependence structure when $q = 1$ or 2, coupled with or without the Box-Cox transformation. The criterion used in the comparison is MARD. For example, for the prediction of V_7, the best prediction result occurs when a common variance is assumed for the AR(2) dependence coupled with the Box-Cox transformation. The resulting MARD is 0.0354, which is slightly better than the result with the above grouping variances. Meanwhile, the best model for the prediction of $V_6 - V_7$, $V_5 - V_7$, and $V_4 - V_7$ is the untransformed AR(1) dependence, having quadratic growth function and with the grouping variances $\{(1)\,(2)\,(3)\,(4)\,(5,6,7)\}$. Thus, with the exception of predicting V_7 alone, the best model seems to be the untransformed AR(1) dependence, having quadratic growth function and with the grouping variances $\{(1)\,(2)\,(3)\,(4)\,(5,6,7)\}$. Of course, the finding is consistent with the variance estimates as shown in Figure 2.

Table 3

The MARD between the predicted and the actual values of $\mathbf{V}^{(2)}$ with the Box-Cox transformation and linear growth function: mice data

Covariance structure	V_7	$V_6 - V_7$	$V_5 - V_7$	$V_4 - V_7$
$\{(1)\,(2)\,(3)\,(4)\,(5,6,7)\}$ & AR(1)	0.0391	0.0549	0.0610	0.0748
$\{(1,2,3,4,5,6,7)\}$ & AR(1)	0.0418	0.0729	0.0798	0.1067
$\{(1)\,(2)\,(3)\,(4)\,(5,6,7)\}$ & AR(2)	0.0367	0.0638	0.0654	0.0847
$\{(1,2,3,4,5,6,7)\}$ & AR(2)	0.0354	0.0836	0.0777	0.1084

4.2 *The drug dissolution data*

The data given in Lee et al. (1999) are the dissolution rates of three standard lots and one test lot. For each lot, there are twelve tablets and for each tablet the dissolution rates are measured at seven time points: 1, 2, 3, 4, 6, 8 and 10 minutes. We used the pooled data of three lots and removed the observations at time 1 and 3 to create an equally-spaced dataset. The dissolution function $F(t)$ of a drug is defined to be the percentage of a tablet that has dissolved at time t, and $R(t)$ is defined by $F(t)/(100 - F(t))$. The data transformation is plotted in Figure 3. Since $0 \leq F(t) \leq 100$, and $R(t) \geq 0$, both ranges are not the entire real line. It may cause the out of range problem when we model them directly. We will therefore consider applying the Box-Cox transformations to $R(t)$ which will avoid the above problem. We observed that there is one possible change point at time 8. Thus, the design matrix $\mathbf{X}$ for the growth curve model is

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 6 & 8 & 8 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}'.$$

From Figure 3, it is clear that $F/(100 - F)$ is not quite linear, but piecewise linear, in time. The Box-Cox transformation applied to $F/(100 - F)$ will help achieve the linearity property. If the time is cut into two pieces, 2–6, 8–10, just as in the design matrix and the Box-Cox transformation is applied to the two pieces separately, the variances are more stable. When different Box-Cox transformations are applied to observations from two different segments, the estimates of $\sigma_1, \ldots, \sigma_5$ for the drug dissolution data are plotted with confidence intervals in Figure 4.
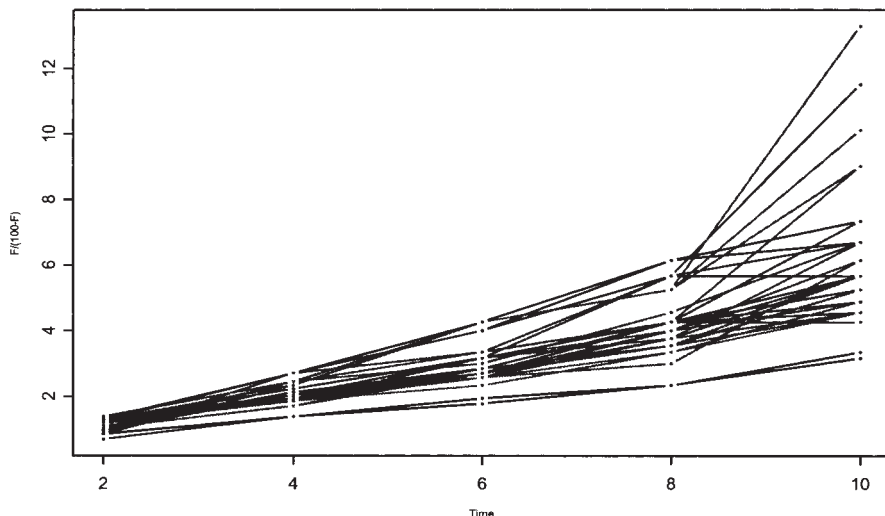

Fig. 3. Plots of dissolution ratio curves

Table 4

The MARD between the predicted and the actual values of $\mathbf{V}^{(2)}$ with the Box-Cox transformation: drug data

Drug data	Covariance structure	V_5	$V_4 - V_5$
Pooled	$\{(1)\,(2,3)\,(4)\,(5)\}$ & AR(1)	0.02672	0.02369
	$\{(1)\,(2)\,(3)\,(4)\,(5)\}$ & AR(1)	0.02689	0.02436
	$\{(1,2,3,4,5)\}$ & AR(1)	0.02742	0.02713

From Figure 4, we found that the standard deviation of the pooled data varies with time. Hence, a possible model is to treat $\sigma_2 = \sigma_3$ and the rest as three different variances in an AR(1) dependence. The value of the log-likelihood function increased from -134.35 under the equality of variances to -113.88 under this grouping variances structure, resulting in a likelihood ratio statistic $Q = 40.94$, with p-value < 0.0001, indicating that the homogeneity of variances is not acceptable. When all 5 variances are distinct, the value of the log-likelihood function is -111.98, resulting in a likelihood ratio statistic $Q^* = 3.80$ and p-value $= 0.0513$, indicating the grouping of variance is acceptable. From Table 4, we found that the model with $q = 1$ and two different variance groups is better than that with equal variance.

4.3 Simulation

In this subsection we will present a simulation study regarding the grouping variances structure. In order to compare different grouping variances, we set
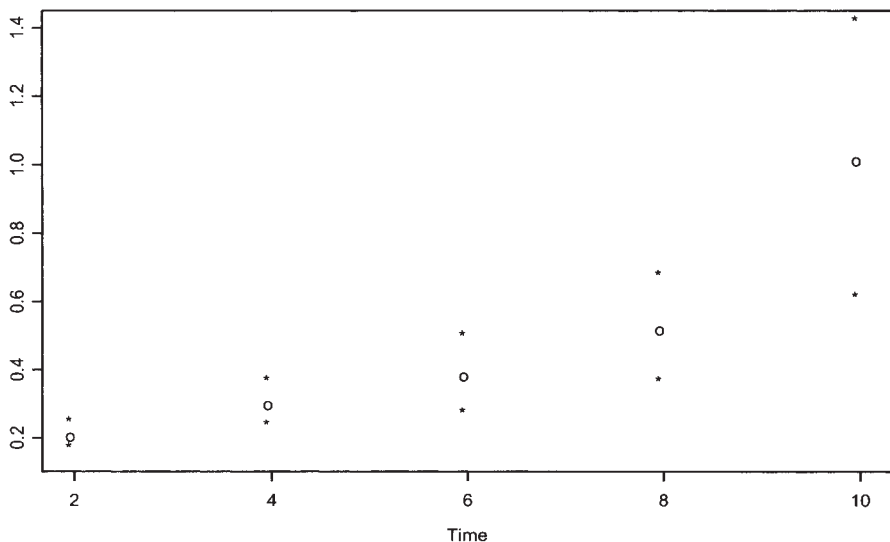

Fig. 4. The estimates of $\sigma_1, \ldots, \sigma_5$ of drug dissolution data with 95% confidence interval for each time point

$\boldsymbol{\tau} = (0.1, 0.2)'$, the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{pmatrix}',$$

$\phi = 0.93$, and the number of replications is 1000. We generate data from grouping variances, $(\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 0.007$ and $\sigma_6 = \sigma_7 = 0.014)$ and $(\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 0.007$ and $\sigma_6 = \sigma_7 = 0.021)$, and present the conditional predictive performance in Table 5 and Table 6, respectively. From the two tables, it is clear that the grouping variances structure yield better and more stable prediction results, as expected.

Table 5

Prediction performance for data generated from grouping variances $(\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 0.007$ and $\sigma_6 = \sigma_7 = 0.014)$ with AR(1) dependence: simulated data

MARD	N	V_7	$V_6 - V_7$	$V_5 - V_7$	$V_4 - V_7$
Grouping variances	12	0.00270	0.00352	0.00358	0.00367
	36	0.00274	0.00355	0.00356	0.00375
Identical variance	12	0.00273	0.00518	0.00449	0.00430
	36	0.00276	0.00516	0.00443	0.00431

Table 6

Prediction performance for data generated from grouping variances $(\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 0.007$ and $\sigma_6 = \sigma_7 = 0.021)$ with AR(1) dependence: simulated data

MARD	N	V_7	$V_6 - V_7$	$V_5 - V_7$	$V_4 - V_7$
Grouping variances	12	0.00414	0.00536	0.00522	0.00502
	36	0.00407	0.00532	0.00511	0.00503
Identical variance	12	0.00428	0.00922	0.00741	0.00674
	36	0.00416	0.00925	0.00723	0.00646

5. Concluding remarks

The model with grouping variances in AR(q) dependence structure provides an effective and practical means of dealing with the growth curve data. An appropriate grouping among the variances in an AR(q) dependence structure gives better and more stable predictive performance. If the growth function is not linear, we can apply the Box-Cox transformation so that the function is linear. If there are change points, we can apply different Box-Cox transformations for different segments of observations so that the function is piecewise linear. If the variance varies over time we can group the variances appropriately in an AR(q) covariance structure. We can plot confidence intervals for standard deviations for each time in

order to observe the adequate grouping and use the likelihood ratio test for the adequacy of grouping the variances. The final grouping of variances is obtained by the prediction results of the last few rows of **Y**, the observed matrix. When the length of the time points is short and the sample size is small, usually $q = 1$ or 2 will suffice. It is conceivable that the grouping of variances will not be a good idea if the variances are somewhat similar. However, if the magnitudes are distinct and in clear clusters, the grouping of variances will certainly improve the predictive performance. Hence, the procedure developed in this paper will be useful for dealing with growth curve data.

Appendix

In this appendix, we will derive the information matrix. $\mathbf{H}_{22}$, $\mathbf{H}_{23}$, $\mathbf{H}_{33}$, $\mathbf{H}_{12}$, $\mathbf{H}_{13}$, $\mathbf{H}_{14}$, $\mathbf{H}_{24}$ and $\mathbf{H}_{34}$ can be obtained from the following equations:

$$
\frac{\partial^2 l}{\partial \sigma_k \, \partial \sigma_j} =
\begin{cases}
-\dfrac{1}{2} \sigma_j^{-2} \sigma_k^{-2} \sum\limits_{i=1}^{N} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau})' \, [\mathbf{E}_{kk} \mathbf{C}^{-1} \mathbf{E}_{jj} \\
\quad + \mathbf{E}_{jj} \mathbf{C}^{-1} \mathbf{E}_{kk}] (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}), & \text{if } k \neq j, \\[2mm]
N \sigma_j^{-2} - \sigma_j^{-3} \sum\limits_{i=1}^{N} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau})' [\mathbf{D}_\sigma^{-1} \mathbf{C}^{-1} \mathbf{E}_{jj} \\
\quad + \mathbf{E}_{jj} \mathbf{C}^{-1} \mathbf{D}_\sigma^{-1} + \sigma_j^{-1} \mathbf{E}_{jj} \mathbf{C}^{-1} \mathbf{E}_{jj}] \, (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}), & \text{if } k = j,
\end{cases}
$$

$$
\frac{\partial^2 l}{\partial \sigma_j \, \partial \phi_k} = -\frac{1}{2} \sigma_j^{-2} \sum_{i=1}^{N} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau})' \, [\mathbf{D}_\sigma^{-1} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \phi_k} \mathbf{C}^{-1} \mathbf{E}_{jj}
$$
$$
+ \mathbf{E}_{jj} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \phi_k} \mathbf{C}^{-1} \mathbf{D}_\sigma^{-1}] \, (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}),
$$

$$
\frac{\partial^2 l}{\partial \phi_j \, \partial \phi_k} = -\frac{N}{2} \operatorname{tr} \left[\mathbf{C}^{-1} \frac{\partial^2 \mathbf{C}}{\partial \phi_j \, \partial \phi_k} - \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \phi_j} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \phi_k} \right]
$$
$$
-\frac{1}{2} \sum_{i=1}^{N} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau})' \, \mathbf{D}_\sigma^{-1} \, \mathbf{C}^{-1}
$$
$$
\times \left[\frac{\partial^2 \mathbf{C}}{\partial \phi_j \, \partial \phi_k} - \frac{\partial \mathbf{C}}{\partial \phi_j} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \phi_k} - \frac{\partial \mathbf{C}}{\partial \phi_k} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \phi_j} \right] \mathbf{C}^{-1} \mathbf{D}_\sigma^{-1} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}),
$$

$$
\frac{\partial^2 l}{\partial \boldsymbol{\tau} \, \partial \sigma_j} = \sigma_j^{-2} \sum_{i=1}^{N} \mathbf{X}' (\mathbf{D}_\sigma^{-1} \, \mathbf{C}^{-1} \mathbf{E}_{jj} + \mathbf{E}_{jj} \mathbf{C}^{-1} \mathbf{D}_\sigma^{-1}) \, (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}),
$$

$$
\frac{\partial^2 l}{\partial \boldsymbol{\tau} \, \partial \phi_j} = \sum_{i=1}^{N} \mathbf{X}' \mathbf{D}_\sigma^{-1} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \phi_j} \mathbf{C}^{-1} \mathbf{D}_\sigma^{-1} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau}),
$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\tau} \, \partial \lambda} = -\sum_{i=1}^{N} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial \mathbf{Y}_i^{(\lambda)}}{\partial \lambda},$$

$$\frac{\partial^2 l}{\partial \sigma_j \partial \lambda} = \sigma_j^{-2} \sum_{i=1}^{N} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau})' [\mathbf{D}_\sigma^{-1}\mathbf{C}^{-1}\mathbf{E}_{jj} + \mathbf{E}_{jj}\mathbf{C}^{-1}\mathbf{D}_\sigma^{-1}]\frac{\partial \mathbf{Y}_i^{(\lambda)}}{\partial \lambda},$$

$$\frac{\partial^2 l}{\partial \phi_j \, \partial \lambda} = \sum_{i=1}^{N} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau})' \mathbf{D}_\sigma^{-1}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \phi_j}\mathbf{C}^{-1}\mathbf{D}_\sigma^{-1}\frac{\partial \mathbf{Y}_i^{(\lambda)}}{\partial \lambda},$$

$$\frac{\partial^2 l}{\partial \lambda^2} = -\sum_{i=1}^{N} \frac{\partial \mathbf{Y}_i^{(\lambda)'}}{\partial \lambda}\boldsymbol{\Sigma}^{-1}\frac{\partial \mathbf{Y}_i^{(\lambda)}}{\partial \lambda} - \sum_{i=1}^{N} (\mathbf{Y}_i^{(\lambda)} - \mathbf{X}\boldsymbol{\tau})'\boldsymbol{\Sigma}^{-1}\frac{\partial^2 \mathbf{Y}_i^{(\lambda)}}{\partial \lambda^2},$$

where $\mathbf{E}_{jj}$ is a $p \times p$ matrix with "1" on the diagonal from $\sum_{i=1}^{j-1} g_j + 1$ to $\sum_{i=1}^{j} g_i$ and zero elsewhere. Since $\mathbf{C}$ is a function of $\boldsymbol{\rho}$, the first two derivatives of $\mathbf{C}$ with respect to $\boldsymbol{\phi}$ can be obtained from the first two derivatives of $\boldsymbol{\rho}$ as given below:

$$\frac{\partial \boldsymbol{\rho}}{\partial \phi_j} = \mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial \phi_j}\mathbf{G}^{-1}\boldsymbol{\phi} - \mathbf{G}^{-1}\mathbf{e}_i,$$

$$\frac{\partial^2 \boldsymbol{\rho}}{\partial \phi_j \, \partial \phi_k} = \mathbf{G}^{-1}\frac{\partial^2 \mathbf{G}}{\partial \phi_j \, \partial \phi_k}\mathbf{G}^{-1}\boldsymbol{\phi} + \mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial \phi_j}\mathbf{G}^{-1}\mathbf{e}_k + \mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial \phi_k}\mathbf{G}^{-1}\mathbf{e}_j$$

$$-\mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial \phi_j}\mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial \phi_k}\mathbf{G}^{-1}\boldsymbol{\phi} - \mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial \phi_k}\mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial \phi_j}\mathbf{G}^{-1}\boldsymbol{\phi},$$

where $\mathbf{e}_i$ is the ith column vector of an $q \times q$ identity matrix $\mathbf{I}_q$. And, the first two derivatives of $\mathbf{Y}_i^{(\lambda)}$ with respect to λ can be easily obtained by the following formulas:

$$\frac{\partial \mathbf{Y}_{ij}^{(\lambda)}}{\partial \lambda} = \begin{cases} -\dfrac{(\mathbf{Y}_{ij}+\gamma)^\lambda - 1}{\lambda^2} + \dfrac{(\mathbf{Y}_{ij}+\gamma)^\lambda \log(\mathbf{Y}_{ij}+\gamma)}{\lambda}, & \lambda \neq 0 \\ 0, & \lambda = 0 \end{cases},$$

$$\frac{\partial^2 \mathbf{Y}_{ij}^{(\lambda)}}{\partial \lambda^2} = \begin{cases} \dfrac{2((\mathbf{Y}_{ij}+\gamma)^\lambda - 1)}{\lambda^3} - \dfrac{2(\mathbf{Y}_{ij}+\gamma)^\lambda \log(\mathbf{Y}_{ij}+\gamma)}{\lambda^2} \\ +\dfrac{(\mathbf{Y}_{ij}+\gamma)^\lambda(\log(\mathbf{Y}_{ij}+\gamma))^2}{\lambda}, & \lambda \neq 0 \\ 0, & \lambda = 0 \end{cases}.$$

Acknowledgement

References

FEARN, T., 1975: A Bayesian Approach to Growth Curve. *Biometrika* **62**, 89–100.

GEARY, D. N., 1989: Modeling the Covariance Structure of Repeated Measurements. *Biometrics* **45**, 1183–1195.

GEISSER, S., 1970: Bayesian Estimation in Multivariate Analysis. *Sankhya* **A32**, 53–64.

GEISSER, S., 1980: Growth Curve Analysis. In P. R. Krishnaiah (ed.):*Handbook of Statistics* (Vol. 1). North-Holland, Amsterdam, 89–115.

JENNRICH, R. I. and SCHLUCHTER, M. D., 1986: Unbalanced Repeated-Measures Models with Structured Covariance Matrices. *Biometrics* **42**, 805–820.

KERAMIDAS, E. M. and LEE, J. C., 1990: Forecasting Technological Substitutions with Concurrent Short Time Series. *Journal of the American Statistical Association* **85**, 625–632.

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 MANOVA Model Applied to Problems in Growth Curves. *Annals of the Institute of Statistical Mathematics* **18**, 75–86.

LAIRD, N. M. and WARE, J. H. 1982: Random-Effects Models for Longitudinal Data. *Biometrics* **38**, 963–974.

LEE, J. C., 1988: Prediction and Estimation of Growth Curves with Special Covariance Structures. *Journal of the American Statistical Association* **83**, 432–440.

LEE, J. C., 1991: Test and Model Selection for the General Growth Curve Model. *Biometrics* **47**, 147–159.

LEE, J. C., CHEN, D. T., HUNG, H. N. and CHEN, J. J., 1999: Analysis of Drug Dissolution. *Statistics in Medicine* **18**, 799–814.

LEE, J. C. and GEISSER, S., 1972: Growth Curve Prediction. *Sankhya* **A34**, 393–412.

LEE, J. C. and GEISSER, S., 1975: Applications of Growth Curve Prediction. *Sankhya* **A37**, 239–256.

LEE, J. C. and LU, K. W., 1987: On a Family of Data-based Transformed Models Useful in Forecasting technological Substitutions. *Technological Forecasting and Social Change* **31**, 61–78.

MANSOUR, H., NORDHEIM, E. V. and RUTLEDGE, J. J., 1985: Maximum Likelihood Estimation of Variance Components in Repeated-Measures Designs Assuming Autoregressive Errors. *Biometrics* **41**, 287–294.

PAN, J. X., FANG, K. T., and VON ROSEN, D., 1999: Bayesian Local Influence in Growth Curve Model with Unstructured Covariance. *Biometrical Journal* **41**, 641–658.

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.

RAO, C. R, 1984: Prediction of Future Observations in Polynomial Growth Curve Model. In Application and New Directions: *Proc. Indian Statistical Institute Golden Jubilee International Conference on Statistics.* Indian Statistical Institute, Calcutta, 512–520.

RAO, C. R., 1987: Prediction of Future Observations in Growth Curve Model. *Statistics Science* **2**, 434–471.

ROCHON, J., 1992: ARMA Covariance Structures with Time Heteroscedasticity for Repeated Measures Experiments. *Journal of the American Statistical Association* **87**, 777–784.

VON ROSEN, D., 1991: The Growth Curve Model: A Review. *Communications in Statistics-Theory and Methods* **20**, 2791–2822.

WILLIAMS, J. S. and IZENMAN, A. J., 1981: A Class of Linear Spectral Models and Analysis for the Study of Longitude Data. *Technical Report,* Dept. of Statistics, Colorado State University.