ELSEVIER

Journal of Statistical Planning and Inference 64 (1997) 205–229

journal of statistical planning and inference

Bayesian analysis of growth curves with AR(1) dependence[1]

Jack C. Lee*, Y.L. Hsu

Institute of Statistics, National Chiao Tung University, 1001 TA Hsueh Road, Hsinchu, Taiwan 30050, ROC

Received 19 January 1996; received in revised form 23 December 1996

Abstract

In this paper we consider Bayesian analysis of the generalized growth curve model when the covariance matrix $\Sigma = \sigma^2 C$ where $C = (\rho^{|i-j|})$, $\sigma^2 > 0$ and $-1 < \rho < 1$ are unknown. We consider both parameter estimation and prediction of future values. Results are illustrated with real and simulated data. © 1997 Elsevier Science B.V.

Keywords: Predictions; Noninformative priors; Trace T distribution

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 Rao (1965, 1966, 1967, 1977, 1984, 1987), Khatri (1966, 1973), Grizzle and Allen (1969), Geisser (1970, 1980, 1981), Lee and Geisser (1972, 1975, 1996), Fearn (1975), Lee (1982, 1988c, 1991), Jennrich and Schluchter (1986), among others.

The generalized growth curve model is defined as

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

where τ is unknown and X and A are known design matrices of ranks $m < p$ and $r < N$, respectively. The columns of ε are independent p-variate normal, with mean vector $\mathbf{0}$

[1] Research Supported in Part by NSC grant 84-2121-M009-008. The authors thank A. Chao, a referee and an associate editor for some helpful comments on an earlier version of the paper, K.C. Liu and W.F. Niu for programming support.

* Corresponding author. e-mail: jclee@stat.nctu.edu.tw

PII S0378-3758(96)00144-9

and common covariance matrix Σ. In general, p is the number of time (or spatial) points observed on each of the N cases, m and r, which usually specify the degree of polynomial in time (or space) and the number of distinct groups, respectively, are assumed known. The design matrices X and A will therefore characterize the degree of the growth function and the distinct grouping out of the N independent vector observations. Potthoff and Roy (1964) gave many examples of growth curve applications for the model (1.1). Grizzle and Allen (1969), Lee and Geisser (1975), Rao (1977, 1987) and Lee (1988c), Chi and Reinsel (1989), among others, applied the model to some biological data. Lee (1988a) and Keramidas and Lee (1990) applied the model to the forecast of technology substitutions.

In Lee (1988c, 1991) and Keramidas and Lee (1990, 1995) the importance of the AR(1) dependence, or serial covariance structure, was demonstrated repeatedly for the covariance matrix Σ for the model (1.1). When the AR(1) dependence holds for Σ, we have $\Sigma = \sigma^2 C$, where $C = (\rho^{|i-j|})$, for $i, j = 1, \ldots, p$, $\sigma^2 > 0$ and $-1 < \rho < 1$ are unknown. The estimation of the parameters and prediction of future values for this model have been so far based on the method of maximum likelihood (ML), which is optimum in large sample. The purpose of the paper is to consider this model from a Bayesian point of view hoping that a more practical solution can be furnished when the sample size is relatively small. Indeed, several published data sets are relatively small in their sizes. We will compare our results with those based on the ML method via real and simulated data sets.

The serial covariance structure is defined as

$$\Sigma = \sigma^2 C, \tag{1.2}$$

where $C = (\rho^{|i-j|})$, $i, j = 1, \ldots, p$, i.e.,

$$C = \begin{bmatrix} 1 & \rho & \cdots & \rho^{p-1} \\ \rho & 1 & \cdots & \rho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \cdots & 1 \end{bmatrix} \tag{1.3}$$

and $\sigma^2 > 0$ and $-1 < \rho < 1$ are unknown. It is conceivably one of the most important covariance structures for the generalized growth curve model.

In addition to the inferences on the parameters τ, σ^2 and ρ, we will also consider several types of prediction problem for the growth curve model as specified by (1.1)–(1.3). Let V be a set of $p \times K$ future observations drawn from the generalized growth curve model, i.e. the set of future observations are such that given the parameters τ and Σ,

$$E(V) = X\tau F, \tag{1.4}$$

where $E(\,)$ denotes expected value, F is a known $r \times K$ matrix, and the columns of V are independent and multivariate normal with a common covariance matrix Σ.

Geisser (1970, 1980) and Lee (1982) considered prediction of V, given Y as the sample, from a Bayesian viewpoint. Lee and Geisser (1972, 1975), Fearn (1975), Rao (1975), and Lee (1988c, 1991) considered the problem of predicting $V^{(2)}$, given $V^{(1)}$ and Y, if V is partitioned as $V = (V^{(1)'}, V^{(2)'})'$, where $V^{(i)}$ 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 cases 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.

The third prediction problem is somewhat different. It is concerned with predicting the future values of the observed cases. Let y, of dimension $q \times n$, be a set of n ($\leqslant N$) future q-dimensional observations whose previous p-dimensional observations are a subset of Y. We are interested in predicting y given Y. This is a time series prediction and thus is important in practice. This type of prediction is called the *extended prediction* of y, because the prediction is made beyond the observed time range of the sample Y. The extended prediction of y was considered by Lee (1988c) and Keramidas and Lee (1990, 1995).

In Section 2, Bayesian estimation of the parameters is considered. Section 3 is concerned with three types of prediction problem. The results developed in the paper are illustrated in Section 4 with real and simulated data. Finally, some concluding remarks are given in Section 5.

2. Bayesian estimation of parameters

For the sake of convenience we shall deal with the pseudo-augmented model

$$E(Y) = (X, Z)\begin{pmatrix}\tau \\ \mathbf{0}\end{pmatrix}A. \tag{2.1}$$

The likelihood function of τ, σ^2 and ρ given Y is

$$L(\tau, \sigma^2, \rho|Y) \propto \sigma^{-pN}(1-\rho^2)^{-(p-1)N/2} \times \exp\left\{-\frac{1}{2\sigma^2}\operatorname{tr} C^{-1}\left[Y - (X, Z)\begin{pmatrix}\tau \\ \mathbf{0}\end{pmatrix}A\right]\left[Y - (X, Z)\begin{pmatrix}\tau \\ \mathbf{0}\end{pmatrix}A\right]'\right\}. \tag{2.2}$$

For the prior of τ, σ^2 and ρ, we will use the following noninformative prior

$$g(\tau, \sigma^2, \rho) \propto \frac{1}{\sigma^2}. \tag{2.3}$$

In (2.3), we have assumed that τ, σ^2 and ρ have independent prior distributions and no information is available for each of the parameters. This is in the same spirit as Zellner and Tiao (1964).

Hence, the posterior density of τ, σ^2 and ρ given $\boldsymbol{Y}$ is

$$P(\tau, \sigma^2, \rho|\boldsymbol{Y}) \propto \sigma^{-(pN+2)}(1-\rho^2)^{-(p-1)N/2} \times \exp\left\{-\frac{1}{2\sigma^2}\operatorname{tr}\boldsymbol{C}^{-1}\left[\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{Z})\begin{pmatrix}\tau\\ \boldsymbol{0}\end{pmatrix}\boldsymbol{A}\right]\left[\boldsymbol{Y}-(\boldsymbol{X},\boldsymbol{Z})\begin{pmatrix}\tau\\ \boldsymbol{0}\end{pmatrix}\boldsymbol{A}\right]'\right\}. \tag{2.4}$$

Integrating out σ^2 and using Lemma A.1 and the application of some algebraic identities yield

$$P(\tau, \rho|\boldsymbol{Y}) \propto (1-\rho^2)^{-(p-1)N/2} S_1^{-pN/2}, \tag{2.5}$$

where

$$\begin{aligned} S_1 &= \operatorname{tr}(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})(\tau-\hat{\tau})\boldsymbol{A}\boldsymbol{A}'(\tau-\hat{\tau})' + b,\\ \hat{\tau} &= (\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{Y}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1},\\ b &= \operatorname{tr}(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{X} + \operatorname{tr}(\boldsymbol{Z}'\boldsymbol{C}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}\boldsymbol{Y}'\boldsymbol{Z},\\ \boldsymbol{S} &= \boldsymbol{Y}[\boldsymbol{I}-\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1}\boldsymbol{A}]\boldsymbol{Y}'. \end{aligned} \tag{2.6}$$

By Theorem A.1, it is clear that conditional on ρ,

$$\underset{m\times r}{\tau}\,|\rho \sim \operatorname{Tr}(\hat{\tau}, \boldsymbol{A}\boldsymbol{A}', b, \boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X}, pN). \tag{2.7}$$

Moreover, integrating out τ from (2.5) we have

$$P(\rho|\boldsymbol{Y}) \propto b^{-(pN-mr)/2}|\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X}|^{-r/2}(1-\rho^2)^{-(p-1)N/2}. \tag{2.8}$$

Since $P(\tau, \rho|\boldsymbol{Y}) = P(\tau|\rho, \boldsymbol{Y})P(\rho|\boldsymbol{Y})$ and $P(\tau|\boldsymbol{Y}) = \int P(\tau, \rho|\boldsymbol{Y})\,\mathrm{d}\rho$, the integral can be approximated by

$$P(\tau|\boldsymbol{Y}) \doteq P(\tau|\hat{\rho}, \boldsymbol{Y}), \tag{2.9}$$

where $\hat{\rho}$ is the mode of $P(\rho|\boldsymbol{Y})$, if $P(\rho|\boldsymbol{Y})$ is concentrated and nearly symmetric, as pointed out by Ljung and Box (1980). Of course, the integration can be performed numerically as the one-dimensional integral can be done rather accurately.

Hence, we have the following posterior distribution of τ:

$$P(\tau|\boldsymbol{Y}) \doteq \operatorname{Tr}(\hat{\tau}^*, \boldsymbol{A}\boldsymbol{A}', \hat{b}, \boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X}, pN), \tag{2.10}$$

where

$$\begin{aligned} \hat{\tau}^* &= (\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{Y}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1},\\ \hat{b} &= \operatorname{tr}(\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{S}\hat{\boldsymbol{C}}^{-1}\boldsymbol{X} + \operatorname{tr}(\boldsymbol{Z}'\hat{\boldsymbol{C}}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}\boldsymbol{Y}'\boldsymbol{Z},\\ \hat{\boldsymbol{C}} &= (\hat{\rho}^{|i-j|}), \end{aligned} \tag{2.11}$$

$\hat{\rho}$ maximizes $P(\rho|\boldsymbol{Y})$, as given in (2.8).

Thus, a posteriori, τ has a *trace T* distribution.

Integrating w.r.t. τ in (2.4) and using arguments similar to (2.9), we obtain the following approximation for the posterior distribution of σ^2,

$$P(\sigma^2|\boldsymbol{Y}) \doteq \mathrm{IG}\left(\frac{pN-mr}{2}, \frac{\hat{b}}{2}\right), \tag{2.12}$$

where $\hat{b}$ is given in (2.11) and $\mathrm{IG}(v_1, v_2)$ is the inverse gamma distribution with parameters v_1 and v_2.

By Theorem A.3 and the fact that if $V \sim \mathrm{Beta}(\frac{1}{2}v_2, \frac{1}{2}v_1)$ and $F \sim F(v_1, v_2)$, then $V = v_2/(v_2 + v_1F)$, $1-\alpha$ posterior region for τ can be obtained from the following inequality:

$$\hat{b}^{-1}\mathrm{tr}(\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})(\tau-\hat{\tau}^*)\boldsymbol{A}\boldsymbol{A}'(\tau-\hat{\tau}^*)' \leqslant \frac{mr}{pN-mr}F_{1-\alpha}(mr, pN-mr) \tag{2.13}$$

where $F_{1-\alpha}(v_1, v_2)$ is the upper 100α percent point of the F-distribution.

For the special case in which $r = 1$, we have the following $1-\alpha$ posterior region for τ:

$$N\hat{b}^{-1}(\tau-\hat{\tau}^*)'(\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})(\tau-\hat{\tau}^*) \leqslant \frac{m}{pN-m}F_{1-\alpha}(m, pN-m), \tag{2.14}$$

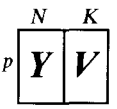
$\hat{b}$ and $\hat{\tau}^*$ are given in (2.11).

3. Prediction

Three types of prediction for the model specified by (1.1)–(1.3) will be considered in this section.

3.1. Prediction of the whole future matrix $\boldsymbol{V}$

The prediction of the future matrix $\boldsymbol{V}$, $p \times K$, given the sample, $\boldsymbol{Y}$, $p \times N$, is considered here. Schematically, the matrices $\boldsymbol{Y}$ and $\boldsymbol{V}$ are shown below.



The density function of $\boldsymbol{V}$ given τ, σ^2 and ρ is

$$f(\boldsymbol{V}|\tau, \sigma^2, \rho) \propto \sigma^{-pK}(1-\rho^2)^{-(p-1)K/2}\exp\left[-\frac{1}{2\sigma^2}\mathrm{tr}\,\boldsymbol{C}^{-1}(\boldsymbol{V}-\boldsymbol{X}\tau\boldsymbol{F})(\boldsymbol{V}-\boldsymbol{X}\tau\boldsymbol{F})'\right]. \tag{3.1}$$

Upon combining with the posterior density of τ, σ^2 and ρ as given in (2.4), and integrating out σ^2, we have

$$P(\boldsymbol{V}, \tau, \rho|\boldsymbol{Y}) \propto (1-\rho^2)^{-(p-1)(N+K)/2} S_2^{-p(N+K)/2}, \tag{3.2}$$

where

$$\begin{aligned} S_2 &= \operatorname{tr} \boldsymbol{C}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\tau\boldsymbol{A})(\boldsymbol{Y}-\boldsymbol{X}\tau\boldsymbol{A})' + \operatorname{tr} \boldsymbol{C}^{-1}(\boldsymbol{V}-\boldsymbol{X}\tau\boldsymbol{F})(\boldsymbol{V}-\boldsymbol{X}\tau\boldsymbol{F})', \\ &= \operatorname{tr} \boldsymbol{C}^{-1}(\boldsymbol{Y}_0-\boldsymbol{X}\tau\boldsymbol{A}_0)(\boldsymbol{Y}_0-\boldsymbol{X}\tau\boldsymbol{A}_0)', \\ \boldsymbol{Y}_0 &= (\boldsymbol{Y} \quad \boldsymbol{V}), \\ \boldsymbol{A}_0 &= (\boldsymbol{A} \quad \boldsymbol{F}). \end{aligned} \tag{3.3}$$

By Lemma A.1, we obtain

$$S_2 = b_0 + \operatorname{tr} \boldsymbol{A}_0\boldsymbol{A}_0'(\tau-\hat{\tau}_0)'(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})(\tau-\hat{\tau}_0), \tag{3.4}$$

where

$$\begin{aligned} b_0 &= \operatorname{tr}(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{S}_0\boldsymbol{C}^{-1}\boldsymbol{X} + \operatorname{tr}(\boldsymbol{Z}'\boldsymbol{C}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}_0\boldsymbol{Y}_0'\boldsymbol{Z}, \\ \boldsymbol{S}_0 &= \boldsymbol{Y}_0[\boldsymbol{I}-\boldsymbol{A}_0'(\boldsymbol{A}_0\boldsymbol{A}_0')^{-1}\boldsymbol{A}_0]\boldsymbol{Y}_0', \\ \hat{\tau}_0 &= (\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{Y}_0\boldsymbol{A}_0'(\boldsymbol{A}_0\boldsymbol{A}_0')^{-1}. \end{aligned} \tag{3.5}$$

Integrating out τ, we have

$$P(\boldsymbol{V}, \rho|\boldsymbol{Y}) \propto b_0^{-(p(N+K)-mr)/2}|\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X}|^{-r/2}(1-\rho^2)^{-(p-1)(N+K)/2}. \tag{3.6}$$

By (4.12) of Geisser (1970),

$$\boldsymbol{S}_0 = \boldsymbol{S} + (\boldsymbol{V}-\boldsymbol{Y}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1}\boldsymbol{F})\boldsymbol{M}(\boldsymbol{V}-\boldsymbol{Y}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1}\boldsymbol{F})', \tag{3.7}$$

where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{F}'(\boldsymbol{A}_0\boldsymbol{A}_0)^{-1}\boldsymbol{F}$, and it can be shown that

$$\begin{aligned} b_0 = b &+ \operatorname{tr} \boldsymbol{M}(\boldsymbol{V}-\boldsymbol{X}\hat{\tau}\boldsymbol{F})'\boldsymbol{C}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}^{-1}(\boldsymbol{V}-\boldsymbol{X}\hat{\tau}\boldsymbol{F}) \\ &+ \operatorname{tr}(\boldsymbol{V}-\boldsymbol{X}\hat{\tau}\boldsymbol{F})'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{C}\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\boldsymbol{V}-\boldsymbol{X}\hat{\tau}\boldsymbol{F}), \end{aligned} \tag{3.8}$$

where $\hat{\tau}$ is given in (2.6).

We will consider the special situation in which $r = 1$. In this situation $\boldsymbol{F} = 1$ and $\boldsymbol{M}$ is a constant. Integration w.r.t. $\boldsymbol{V}$ yields

$$P_V(\rho|\boldsymbol{Y}) \propto b^{-(pN-mr)/2}|\boldsymbol{G}|^{-K/2}|\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X}|^{-r/2}(1-\rho^2)^{-(p-1)(N+K)/2}. \tag{3.9}$$

where

$$\boldsymbol{G} = \boldsymbol{M}\boldsymbol{C}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}^{-1} + \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{C}\boldsymbol{Z})^{-1}\boldsymbol{Z}' \tag{3.10}$$

With arguments similar to (2.9), we obtain the following approximation for the predictive distribution of V:

$$P(V|Y) \doteq \mathrm{Tr}(X\hat{\tau}_V F, 1, \hat{b}_V, \hat{G}_V, p(N+K)-mr), \tag{3.11}$$

where

$$\begin{aligned}
\hat{\tau}_V &= (X'\hat{C}_V^{-1}X)^{-1}X'\hat{C}_V^{-1}YA'(AA')^{-1},\\
\hat{b}_V &= \mathrm{tr}(X'\hat{C}_V^{-1}X)^{-1}X'\hat{C}_V^{-1}S\hat{C}_V^{-1}X + \mathrm{tr}(Z'\hat{C}_V Z)^{-1}Z'YY'Z,\\
\hat{G}_V &= M\hat{C}_V^{-1}X(X'\hat{C}_V^{-1}X)^{-1}X'\hat{C}_V^{-1} + Z(Z'\hat{C}_V Z)^{-1}Z',\\
\hat{C}_V &= (\hat{\rho}_V^{|i-j|}),
\end{aligned} \tag{3.12}$$

$\hat{\rho}_V$ maximizes $P_V(\rho|Y)$ as given in (3.9).
It is noted that $\hat{\tau}_V$ and $\hat{\tau}$ are different because (3.9) is different from (2.8). Thus, the future matrix V has a *trace T* distribution.

3.2. Conditional prediction of $V^{(2)}$ given $V^{(1)}$ and Y

We next consider the conditional prediction of $V^{(2)}$ given $V^{(1)}$ and Y, if V is partitioned as $V = (V^{(1)'}, V^{(2)'})'$. Schematically, we have the following:

Y	$V^{(1)}$
	$V^{(2)}$

where Y, $p \times N$, is the complete sample; $V^{(1)}$, $p_1 \times K$, is the partially observed matrix and $V^{(2)}$, $p_2 \times K$, is the unobserved portion to be predicted. Of course, $p_1 + p_2 = p$. From Eq. (3.6), it can be shown that

$$\begin{aligned}
P(V^{(2)}, \rho|V^{(1)}, Y) \propto\ & (1-\rho^2)^{-(p-1)(N+K)/2}|X'C^{-1}X|^{-r/2}\\
& \times [b_1 + (V^{(2)} - \hat{V}^{(2)})'G_{22}(V^{(2)} - \hat{V}^{(2)})]^{-(p(N+K)-mr)/2},
\end{aligned} \tag{3.13}$$

where

$$\begin{aligned}
G &= MC^{-1}X(X'C^{-1}X)^{-1}X'C^{-1} + Z(Z'CZ)^{-1}Z',\\
b_1 &= b + (V^{(1)} - X^{(1)}\hat{\tau}F)'G_{11.2}(V^{(1)} - X^{(1)}\hat{\tau}F),\\
\hat{V}^{(2)} &= X^{(2)}\hat{\tau}F - G_{22}^{-1}G_{21}(V^{(1)} - X^{(1)}\hat{\tau}F),\\
X &= (X^{(1)'}, X^{(2)'})',\\
G &= \begin{pmatrix} G_{11} & G_{21} \\ G_{12} & G_{22} \end{pmatrix}, \quad G_{ij} \text{ is } p_i \times p_j,\\
G_{11.2} &= G_{11} - G_{12}G_{22}^{-1}G_{21}.
\end{aligned} \tag{3.14}$$

Integrating over $V^{(2)}$, we have

$$P(\rho|V^{(1)}, Y) \propto (1-\rho^2)^{-(p-1)(N+K)/2}|X'C^{-1}X|^{-r/2}b_1^{-(pN+p_1K-mr)/2}|G_{22}|^{-p_2/2}. \quad (3.15)$$

With arguments similar to (2.9), we obtain the following approximation for the conditional distribution of $V^{(2)}$ given $V^{(1)}$ and Y:

$$P(V^{(2)}|V^{(1)}, Y) \doteq \mathrm{Tr}(\hat{V}^{(2)^*}, 1, \hat{b}_1, \hat{G}_{22}, p(N+K)-mr), \quad (3.16)$$

where

$$\hat{V}^{(2)^*} = X^{(2)}\hat{\tau}_1 F - \hat{G}_{22}^{-1}\hat{G}_{21}(V^{(1)} - X^{(1)}\hat{\tau}_1 F),$$

$$\hat{b}_1 = \hat{b} + (V^{(1)} - X^{(1)}\hat{\tau}_1 F)'\hat{G}_{11.2}(V^{(1)} - X^{(1)}\hat{\tau}_1 F),$$

$$\hat{\tau}_1 = (X'\hat{C}^{*^{-1}}X)^{-1}X'\hat{C}^{*^{-1}}YA'(AA')^{-1},$$

$$\hat{G} = M\hat{C}^{*^{-1}}X(X'\hat{C}^{*^{-1}}X)^{-1}X'\hat{C}^{*^{-1}} + Z(Z'\hat{C}^*Z)^{-1}Z', \quad (3.17)$$

$\hat{G}_{ij}$ is defined as G_{ij},

$\hat{C}^* = (\hat{\rho}_1^{|i-j|})$,

and $\hat{\rho}_1$ maximizes $P(\rho|V^{(1)}, Y)$ as given in (3.15).

Thus, the predictive inference on $V^{(2)}$ given $V^{(1)}$ and Y can be based on the *trace* T distribution. Alternatively, $P(V^{(2)}|V^{(1)}, Y)$ can be obtained from (3.13) by integrating over ρ numerically, i.e., $P(V^{(2)}|V^{(1)}, Y) = \int P(V^{(2)}, \rho|V^{(1)}, Y)\,\mathrm{d}\rho$.

3.3. *Extended prediction of y*

We now consider the extended prediction of y, given Y. This is a time-series prediction which is of practical interest for many types of growth curve data. In order to make this type of prediction the covariance structure generally has to be extendable to the future values of the individuals observed.

Let x, $q \times m$, be a design matrix corresponding to y, $Y = (Y_1, \ldots, Y_N)$, $A = (A_1, \ldots, A_N)$, $y = (y_1, \ldots, y_n)$, and assume that for $i \leqslant n$,

$$E\begin{pmatrix} Y_i \\ y_i \end{pmatrix} = \begin{pmatrix} X \\ x \end{pmatrix}\tau A_i \quad (3.18)$$

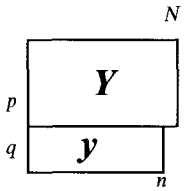
and

$$\Sigma = \mathrm{Cov}\begin{pmatrix} Y_i \\ y_i \end{pmatrix} = \sigma^2\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \quad (3.19)$$

where

$$\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = (\rho^{|a-b|}),$$

$a, b = 1, \ldots, (p+q)$, C_{11} is $p \times p$, C_{12} is $p \times q$, C_{22} is $q \times q$, and $C_{21} = C_{12}'$. Schematically, Y and y are shown below:



For the extended prediction of y, given Y, we will consider the special situation in which $r = n = 1$. This is good enough, in practice, because conditional on the knowledge of τ and $\Sigma, y_1, \ldots, y_n$ are independent. We will consider the situation in which Y_i will be excluded from the sample when y_i is being predicted. Thus, in this study the sample is $Y_{(i)}$ which is $Y_{(i)} = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_N)$ and let $Y_i^* = (Y_i', y_i')'$, $X^* = (X', x')'$ and $A_{(i)} = (A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_N)$.

Combining the density of Y_i^* with the posterior of τ, σ^2 and ρ given $Y_{(i)}$ and integrating out σ^2, we have

$$P(Y_i^*, \tau, \rho | Y_{(i)}) \propto (1-\rho^2)^{-((p-1)N+q)/2} S_3^{-(pN+q)/2}, \tag{3.20}$$

where

$$\begin{aligned}
S_3 &= b_2 + (\tau - \hat{\tau})' Q (\tau - \hat{\tau}), \\
b_2 &= (\hat{\tau}_1 - \hat{\tau}_2)' Q_1 Q^{-1} Q_2 (\hat{\tau}_1 - \hat{\tau}_2) + \operatorname{tr}(Z^{*'} C Z^*)^{-1} Z^{*'} Y_i^* Y_i^{*'} Z^* \\
&\quad + \operatorname{tr}(Z' C_{11} Z)^{-1} Z' Y_{(i)} Y_{(i)}' Z + \operatorname{tr}(X' C_{11}^{-1} X)^{-1} X' C_{11}^{-1} S_{(i)} C_{11}^{-1} X, \\
Q_1 &= (N-1) X' C_{11}^{-1} X, \\
Q_2 &= X^{*'} C^{-1} X^*, \\
Q &= Q_1 + Q_2, \\
S_{(i)} &= Y_{(i)} [I - A_{(i)}' (A_{(i)} A_{(i)}')^{-1} A_{(i)}] Y_{(i)}', \\
\hat{\tau}_1 &= (X' C_{11}^{-1} X)^{-1} X' C_{11}^{-1} Y_{(i)} A_{(i)}' (A_{(i)} A_{(i)}')^{-1}, \\
\hat{\tau}_2 &= (X^{*'} C^{-1} X^*)^{-1} X^{*'} C^{-1} Y_i^* A_i' (A_i A_i')^{-1}, \\
\hat{\tau} &= Q^{-1} (Q_1 \hat{\tau}_1 + Q_2 \hat{\tau}_2),
\end{aligned} \tag{3.21}$$

Z^* is $(p+q) \times (p+q-m)$ and of rank $p+q-m$ such that $X^{*'} Z^* = \mathbf{0}$.

Integrating over τ yields

$$P(\rho, \boldsymbol{Y}_i^* | \boldsymbol{Y}_{(i)}) \propto (1-\rho^2)^{-((p-1)N+q)/2} |\boldsymbol{Q}|^{-1/2} b_2^{-(p(N-1)+q-mr)/2}, \tag{3.22}$$

where

$$\begin{aligned}
&\boldsymbol{G}^* = \boldsymbol{C}^{-1}\boldsymbol{X}^*(\boldsymbol{X}^{*'}\boldsymbol{C}^{-1}\boldsymbol{X}^*)^{-1}\boldsymbol{Q}_1\boldsymbol{Q}^{-1}\boldsymbol{Q}_2(\boldsymbol{X}^{*'}\boldsymbol{C}^{-1}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*'}\boldsymbol{C}^{-1} + \boldsymbol{Z}^*(\boldsymbol{Z}^{*'}\boldsymbol{C}\boldsymbol{Z}^*)^{-1}\boldsymbol{Z}^{*'},\\
&\boldsymbol{G}^* = \begin{pmatrix} \boldsymbol{G}_{11}^* & \boldsymbol{G}_{12}^* \\ \boldsymbol{G}_{21}^* & \boldsymbol{G}_{22}^* \end{pmatrix},\\
&\boldsymbol{G}_{11.2}^* = \boldsymbol{G}_{11}^* - \boldsymbol{G}_{12}^*\boldsymbol{G}_{22}^{*-1}\boldsymbol{G}_{21}^*,\\
&b_2 = b_{(i)} + (\boldsymbol{Y}_i - \boldsymbol{X}\hat{\boldsymbol{\tau}}_1\boldsymbol{A}_i)'\boldsymbol{G}_{11.2}^*(\boldsymbol{Y}_i - \boldsymbol{X}\hat{\boldsymbol{\tau}}_1\boldsymbol{A}_i) + (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)'\boldsymbol{G}_{22}^*(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i),\\
&b_{(i)} = \operatorname{tr}(\boldsymbol{X}'\boldsymbol{C}_{11}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}_{11}^{-1}\boldsymbol{S}_{(i)}\boldsymbol{C}_{11}^{-1}\boldsymbol{X} + \operatorname{tr}(\boldsymbol{Z}'\boldsymbol{C}_{11}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}_{(i)}\boldsymbol{Y}_{(i)}'\boldsymbol{Z},\\
&\hat{\boldsymbol{y}}_i = \boldsymbol{x}\hat{\boldsymbol{\tau}}_1\boldsymbol{A}_i - \boldsymbol{G}_{22}^{*-1}\boldsymbol{G}_{21}^*(\boldsymbol{Y}_i - \boldsymbol{X}\hat{\boldsymbol{\tau}}_1\boldsymbol{A}_i).
\end{aligned} \tag{3.23}$$

By arguments similar to (2.9), we have the following approximate predictive density of $\boldsymbol{y}_i$ given $\boldsymbol{Y}$,

$$P(\boldsymbol{y}_i|\boldsymbol{Y}) \doteq \operatorname{Tr}(\hat{\boldsymbol{y}}_i^*, 1, \hat{b}_{(i)} + (\boldsymbol{Y}_i - \boldsymbol{X}\hat{\boldsymbol{\tau}}_1^*)'\hat{\boldsymbol{G}}_{11.2}^*(\boldsymbol{Y}_i - \boldsymbol{X}\hat{\boldsymbol{\tau}}_1^*), \hat{\boldsymbol{G}}_{22}^*, p(N-1)+q-mr), \tag{3.24}$$

where

$$\begin{aligned}
&\hat{\boldsymbol{y}}_i^* = \boldsymbol{x}\hat{\boldsymbol{\tau}}_1^*\boldsymbol{A}_i - \hat{\boldsymbol{G}}_{22}^{*-1}\hat{\boldsymbol{G}}_{21}^*(\boldsymbol{Y}_i - \boldsymbol{X}\hat{\boldsymbol{\tau}}_1^*\boldsymbol{A}_i),\\
&\hat{\boldsymbol{\tau}}_1^* = (\boldsymbol{X}'\hat{\boldsymbol{C}}_{11}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{C}}_{11}^{-1}\boldsymbol{Y}_{(i)}\boldsymbol{A}_{(i)}'(\boldsymbol{A}_{(i)}\boldsymbol{A}_{(i)}')^{-1},\\
&\hat{\boldsymbol{G}}^* = \hat{\boldsymbol{C}}^{-1}\boldsymbol{X}^*(\boldsymbol{X}^{*'}\hat{\boldsymbol{C}}^{-1}\boldsymbol{X}^*)^{-1}\hat{\boldsymbol{Q}}_1\hat{\boldsymbol{Q}}^{-1}\hat{\boldsymbol{Q}}_2(\boldsymbol{X}^{*'}\hat{\boldsymbol{C}}^{-1}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*'}\hat{\boldsymbol{C}}^{-1}\\
&\qquad + \boldsymbol{Z}^*(\boldsymbol{Z}^{*'}\hat{\boldsymbol{C}}\boldsymbol{Z}^*)^{-1}\boldsymbol{Z}^{*'},\\
&\hat{b}_{(i)} = \operatorname{tr}(\boldsymbol{X}'\hat{\boldsymbol{C}}_{11}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{C}}_{11}^{-1}\boldsymbol{S}_{(i)}\boldsymbol{C}_{11}^{-1}\boldsymbol{X} + \operatorname{tr}(\boldsymbol{Z}'\hat{\boldsymbol{C}}_{11}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}_{(i)}\boldsymbol{Y}_{(i)}'\boldsymbol{Z},\\
&\hat{\boldsymbol{C}} = (\hat{\rho}_y^{|i-j|}),\\
&\hat{\rho}_y \text{ maximizes}\\
&P_y(\rho|\boldsymbol{Y}) \propto (1-\rho^2)^{-(p-1)(N-1)/2}|\boldsymbol{C}|^{-1/2}|\boldsymbol{G}_{22}^*|^{-1/2}|\boldsymbol{Q}|^{-1/2}[b_{(i)} + (\boldsymbol{Y}_i - \boldsymbol{X}\hat{\boldsymbol{\tau}}_1\boldsymbol{A}_i)'\\
&\qquad \times \boldsymbol{G}_{11.2}^*(\boldsymbol{Y}_i - \boldsymbol{X}\hat{\boldsymbol{\tau}}_1\boldsymbol{A}_i)]^{-(p(N-1)-mr)/2}.
\end{aligned} \tag{3.25, 3.26}$$

4. Numerical results

This section is devoted to the illustration of the conditional prediction of $\boldsymbol{V}^{(2)}$ given $\boldsymbol{V}^{(1)}$ and $\boldsymbol{Y}$, and the extended prediction of $\boldsymbol{y}$ given $\boldsymbol{Y}$. For the conditional prediction, we will follow Lee and Geisser (1975), Fearn (1975) and Lee (1988c) in setting $K = 1$,

$p_2 = 1$ and $p_1 = p - 1$, that is, we will predict the last observation of the partially observed vector. For the extended prediction, we will set $q = 1$ and $n = 1$, that is, we will predict one future component at a time.

4.1. Simulations

In this subsection we will present some simulation results regarding the parameter estimation for τ and ρ, the conditional prediction of $\boldsymbol{V}^{(2)}$ given $\boldsymbol{V}^{(1)}$ and $\boldsymbol{Y}$, and the extended prediction of $\boldsymbol{y}$ given $\boldsymbol{Y}$ for the special situation in which $r = 1$ and $K = 1$. The posterior region for τ can be obtained from Eq. (2.14). Meanwhile, from Lee (1988c, 1991), we can also obtain an approximate confidence region for τ. From the asymptotic result of $\hat{\tau}$, the MLE of τ, we have

$$\operatorname{Cov}(\hat{\tau}) \doteq \sigma^2(\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})^{-1} \otimes (\boldsymbol{A}\boldsymbol{A}')^{-1},$$

$$\frac{\operatorname{tr}(\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})(\tau - \hat{\tau})\boldsymbol{A}\boldsymbol{A}'(\tau - \hat{\tau})'}{\hat{\sigma}^2} \sim \chi^2_{mr} \quad \text{as } N \to \infty. \tag{4.1}$$

Hence,

$$\Pr\left[\frac{\operatorname{tr}(\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})(\tau - \hat{\tau})\boldsymbol{A}\boldsymbol{A}'(\tau - \hat{\tau})'}{\hat{\sigma}^2} \leqslant \chi^2_{mr}(\alpha)\right] = 1 - \alpha. \tag{4.2}$$

Since $\boldsymbol{A}\boldsymbol{A}' = N$, we have the following $1 - \alpha$ confidence region:

$$(\tau - \hat{\tau})'(\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})(\tau - \hat{\tau}) \leqslant c_2, \tag{4.3}$$

where

$$c_2 = \frac{\hat{\sigma}^2}{N}\chi^2_{mr}(\alpha),$$

$$\hat{\tau} = (\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{Y}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1}.$$

$$\hat{\sigma}^2 = [\operatorname{tr}(\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{C}}^{-1}\boldsymbol{S}\hat{\boldsymbol{C}}^{-1}\boldsymbol{X} + \operatorname{tr}(\boldsymbol{Z}'\hat{\boldsymbol{C}}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}\boldsymbol{Y}'\boldsymbol{Z}]/pN, \tag{4.4}$$

$$\hat{\boldsymbol{C}} = (\hat{\rho}^{|i-j|}),$$

and $\hat{\rho}$ maximizes the profile likelihood function

$$L_{\max}(\rho) = (\hat{\sigma}^2(\rho))^{-pN/2}(1 - \rho^2)^{-N(p-1)/2}. \tag{4.5}$$

In order to compare the regions for τ obtained by the two different methods, the values for N, p, α, τ, σ^2 and ρ are given in Tables 1 and 2. From the tables, it is clear that the percentage of the ML regions covering the true values is consistently smaller than 0.95. The corresponding Bayesian regions are much better, because the percentage is closer to 0.95 for each of the 12 situations. This explains the phenomenon in

Table 1
Comparison of coverage probabilities between approximate confidence region and Bayesian region for τ $(m = 2, r = 1)$

N	ρ	Coverage probability	
		Bayesian	ML
5	0.5	0.944	0.883
5	0.8	0.934	0.872
10	0.5	0.958	0.942
10	0.8	0.918	0.886
15	0.5	0.955	0.938
15	0.8	0.942	0.923

Note: $p = 4$, $1 - \alpha = 0.95$, $\tau = (25, 0.8)'$, $\sigma^2 = 5$ and no. of replications = 1000.

Table 2
Comparison of coverage probabilities between approximate confidence region and Bayesian region for τ $(m = 3, r = 2)$

	ρ	Coverage probability	
		Bayesian	ML
$N_1 = N_2 = 5$	0.5	0.936	0.842
$N_1 = N_2 = 5$	0.8	0.919	0.823
$N_1 = N_2 = 10$	0.5	0.940	0.905
$N_1 = N_2 = 10$	0.8	0.934	0.891
$N_1 = N_2 = 15$	0.5	0.954	0.932
$N_1 = N_2 = 15$	0.8	0.952	0.932

Note: $p = 4$, $1 - \alpha = 0.95$, $\tau = \begin{pmatrix} 22.3 & 0.27 & 0.26 \\ 20.2 & 1.03 & -0.01 \end{pmatrix}'$, $\sigma^2 = 5$ and no. of replications = 1000.

Fig. 1 in which the confidence region for τ is smaller than the posterior region. This also indicates that the asymptotic test based on the likelihood ratio criterion will be biased toward rejecting the null hypothesis. This type of biasedness has been observed in other occasions as well, see e.g. Laitinen (1978) and Lee (1988b).

We next compare the conditional prediction of $\boldsymbol{Y}^{(2)}$ given $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}$ and the extended prediction of $\boldsymbol{y}$ given $\boldsymbol{Y}$. Here we set $p = 4$, $\alpha = 0.05$, $\tau = (25, 0.8)'$, $\rho = 0.8$, $\sigma^2 = 5$, $N = 5, 10, 15$ and the number of replications $g = 50$. For the conditional prediction, we will set $K = 1$, $p_1 = 3$ and $p_2 = 1$. For predictive purposes, we withhold one vector and use the rest as the sample for predicting the last component of that vector and repeat this for each of the N observations. This gives N predicted values for

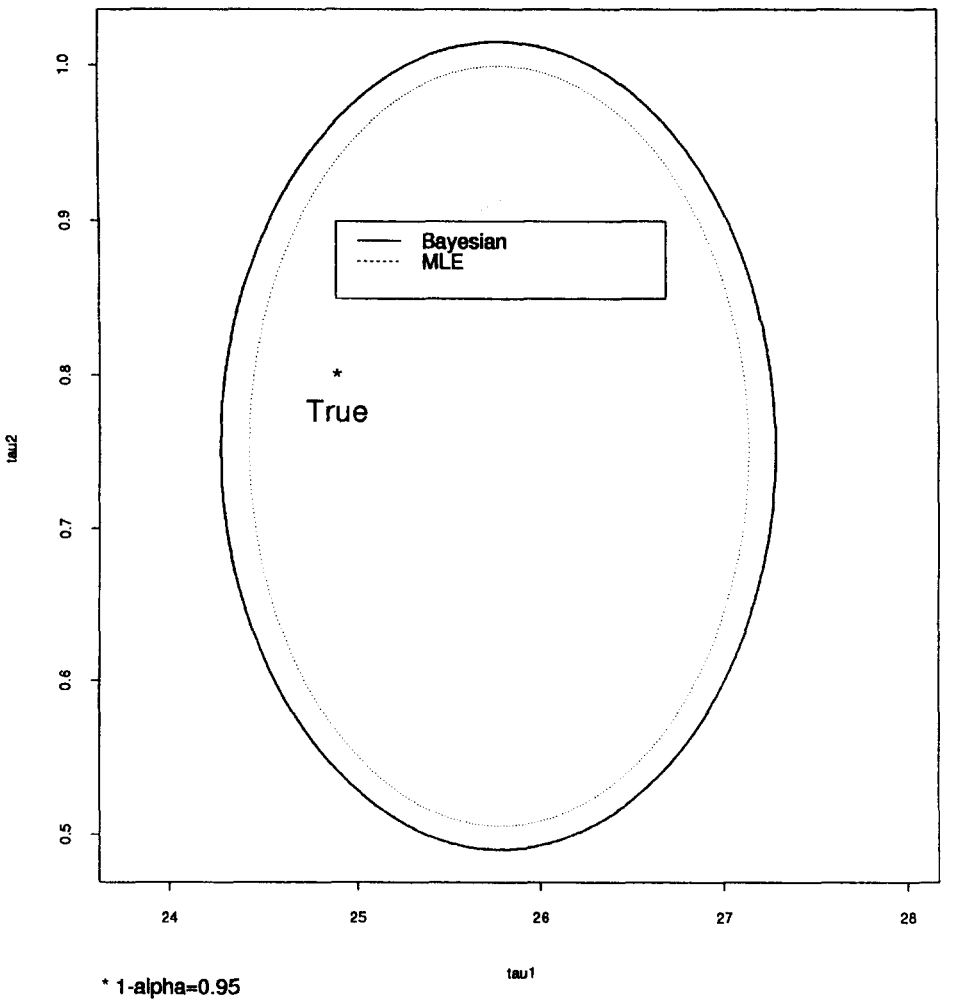


Fig. 1. Confidence and posterior regions for tau.

the last N observed values in each data set. Overall, there are $g \times N$ predicted values to be compared with $g \times N$ actual observations. The mean squared deviation (MSD), the mean absolute deviation (MAD) and the mean relative absolute deviation (MARD) of the predicted values from the actual observations are used to assess the relative merits of the two methods. A comparison of prediction accuracy for $\boldsymbol{V}^{(2)}$ given $\boldsymbol{V}^{(1)}$ and $\boldsymbol{Y}$ between Bayesian and ML methods is given in Table 3. It is noted that in the prediction process, $\boldsymbol{V}^{(1)}$ is not used in deriving the parameter estimation although it is used as part of the conditional mean vector.

As for the extended prediction, we set the last row of the generated $\boldsymbol{Y}$ as $\boldsymbol{y}$ and the first three rows as the corresponding sample. Thus, we set $p = 3$ and $q = 1$. Similar to

Table 3
Comparison of prediction accuracy for $V^{(2)}$ between Bayesian and ML methods

	Bayesian			ML		
	$N=5$	$N=10$	$N=15$	$N=5$	$N=10$	$N=15$
MSD	2.1504	1.9981	1.8389	2.3752	2.0089	1.8768
MAD	1.1505	1.1329	1.0948	1.2304	1.1423	1.1058
MARD	0.0428	0.0418	0.0402	0.0455	0.0422	0.0407

Note: $p=4$, $1-\alpha=0.95$, $\tau=(25, 0.8)'$, $\rho=0.8$, $\sigma^2=5$ and no. of replications = 50.

Table 4
Comparison of prediction accuracy and coverage probability for y between Bayesian and ML methods

	Bayesian			ML		
	$N=5$	$N=10$	$N=15$	$N=5$	$N=10$	$N=15$
MSD	2.3008	1.9371	1.9048	2.3585	1.9350	1.9066
MAD	1.2106	1.1181	1.1009	1.2341	1.1163	1.1019
MARD	0.0444	0.0412	0.0437	0.0454	0.0415	0.0404
Coverage	0.940	0.949	0.954	0.818	0.902	0.928

Note: $p=3$, $q=1$, $1-\alpha=0.95$, $\tau=(25, 0.8)'$, $\rho=0.8$, $\sigma^2=5$ and no. of replications = 50.

the conditional prediction, there will be $g \times N$ predicted values to be compared with $g \times N$ actuals. A comparison of prediction accuracy for $\boldsymbol{y}$ given $\boldsymbol{Y}$ between Bayesian and ML methods in terms of MSD, MAD and MARD is given in Table 4.

In addition to the point prediction we will also compare the interval prediction for $\boldsymbol{y}$ given $\boldsymbol{Y}$. In order to compare the interval prediction for $\boldsymbol{y}$, we note that the Bayesian method is based on a property of the *trace* T distribution as given in (2.14). For the ML method, we will use the following approximate interval:

$$\hat{\boldsymbol{y}} \pm z_{\alpha/2}\hat{\sigma}_f, \tag{4.6}$$

where $z_{\alpha/2}$ is the $100*\alpha/2$ percent point of the standard normal distribution,

$$\sigma_f^2 = \sigma^2[\boldsymbol{C}_{22} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12} + \boldsymbol{H}(N\boldsymbol{X}'\boldsymbol{C}_{11}^{-1}\boldsymbol{X})^{-1}\boldsymbol{H}' + 2\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{X}(N\boldsymbol{X}'\boldsymbol{C}_{11}^{-1}\boldsymbol{X})^{-1}\boldsymbol{H}'], \tag{4.7}$$

$$\boldsymbol{H} = \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{X} - \boldsymbol{x}. \tag{4.8}$$

It is noted that σ_f^2 is the variance of the forecast error for $\boldsymbol{y}$ when the parameters are assumed known. For the variance of forecast error for $\boldsymbol{V}^{(2)}$ we will use $S_{m_2^*}$ as indicated in (3.11) of Lee (1988c):

$$S_{m_2^*} = \boldsymbol{C}_{22} + (\boldsymbol{b}_1\boldsymbol{B}_2 + \boldsymbol{C}_{21})\boldsymbol{C}_{11}^{-1}(\boldsymbol{b}_1\boldsymbol{B}_2 + \boldsymbol{C}_{21})' - (\boldsymbol{b}_1\boldsymbol{B}_2 + \boldsymbol{C}_{21})\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12},$$
$$- \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}(\boldsymbol{b}_1\boldsymbol{B}_2 + \boldsymbol{C}_{21})', \tag{4.9}$$

$$\boldsymbol{b}_1 = \boldsymbol{F}'(\boldsymbol{F}\boldsymbol{F}')^{-}\boldsymbol{F}, \tag{4.10}$$

$$\boldsymbol{B}_2 = (\boldsymbol{x} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{C}_{11}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'. \tag{4.11}$$

From the simulation study, we see that both methods produce similar prediction accuracy for $\boldsymbol{V}^{(2)}$ and $\boldsymbol{y}$ in terms of MSD, MAD or MARD, as shown in Tables 3 and 4, respectively. The Bayesian method yields longer predictive intervals for $\boldsymbol{y}$, as shown in Fig. 2 for $N = 10$, which is typical in each of the 50 replications for each N which was conducted in the simulation. Similar results are also true with the real data examined in Section 4.2, as shown in Fig. 4. However, the percentage of the Bayesian predictive intervals covering the values to be predicted are closer to $1 - \alpha$ than the ML intervals, as seen in Table 4. This is consistent with the comparison of the regions for τ. It is therefore clear that the Bayesian predictive intervals are superior to the ML intervals when the samples size N is relatively small.

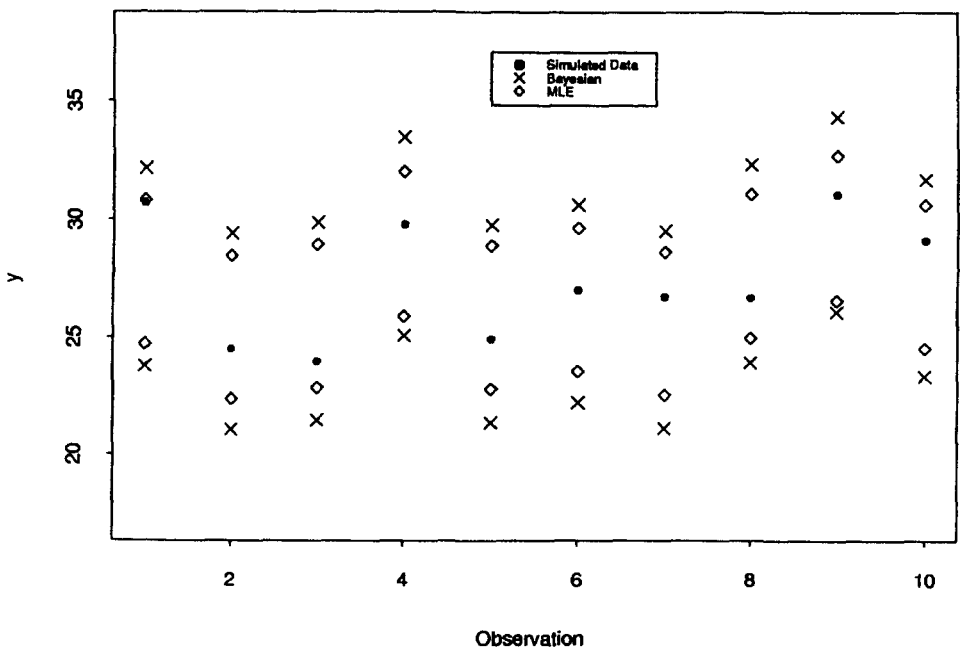


Fig. 2. Comparison of predictive intervals for y for simulated data.

4.2. Illustrative examples

Some of the results developed in Sections 2 and 3 are applied to the dental measurements of 11 girls and 16 boys and three other data sets (ramus, mice and glucose data). The dental data set, which is reproduced in Table 5, was first considered by Potthoff and Roy (1964) and later analyzed by Lee and Geisser (1975), Fearn (1975), Rao (1987), and Lee (1988c, 1991), among others. Dental measurements were made on 11 girls and 16 boys at ages 8, 10, 12, and 14 years. Each measurement is the distance, in millimeters, from the center of the pituitary to the pterygomaxillary fissure. From Table 5 it is clear that the distance being measured can decrease with age, due to the fact that the distance represents the relative position of two points. The ramus data were originally given in Elston and Grizzle (1962) and subsequently analyzed by Lee and Geisser (1975), Fearn (1975), Rao (1987), Lee (1988c, 1991), among others. The mice data were first reported by Williams and Izenman (1981) and later analyzed by Rao (1987), Lee (1988c, 1991), among others. The glucose data were first reported by Zerbe (1979) and later analyzed by Chi and Reinsel (1989) and Keramidas and Lee (1995). We will primarily focus on the dental data, although the prediction results will be summarized for the other three data sets as well.

We will next deal with the dental data in more details. Since the measurements are obtained at equal time intervals, the design matrix X is

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 \end{pmatrix}'.$$

Table 5
Dental measurements of 11 girls and 16 boys

Individual[a]	Age				Individual[a]	Age			
	8	10	12	14		8	10	12	14
1	21	20	21.5	23	15	25.5	27.5	26.5	27
2	21	21.5	24	25.5	16	20	23.5	22.5	26
3	20.5	24	24.5	26	17	24.5	25.5	27	28.5
4	23.5	24.5	25	26.5	18	22	22	24.5	26.5
5	21.5	23	22.5	23.5	19	24	21.5	24.5	25.5
6	20	21	21	22.5	20	23	20.5	31	26
7	21.5	22.5	23	25	21	27.5	28	31	31.5
8	23	23	23.5	24	22	23	23	23.5	25
9	20	21	22	21.5	23	21.5	23.5	24	28
10	16.5	19	19	19.5	24	17	24.5	26	29.5
11	24.5	25	28	28	25	22.5	25.5	25.5	26
12	26	25	29	31	26	23	24.5	26	30
13	21.5	22.5	23	26.5	27	22	21.5	23.5	25
14	23	22.5	24	27.5					

[a] Individuals 1–11 are girls, 12–27 are boys.

Also, from the findings in Lee and Geisser (1975) and Lee (1988c), the individual #20, who is a boy, could be excluded because it is suspected to be an aberrant observation. Furthermore, from Lee (1988c, 1991) it is clear that the data should be treated as from two different populations with distinct mean functions and serial covariance matrices. However, for illustrative purposes, we will also include the situation in which a common serial covariance structure is assumed for measurements of both girls and boys.

Before dwelling on the prediction results we note that from Fig. 3, the posterior densities of ρ for different subsets of the data are all well concentrated and nearly

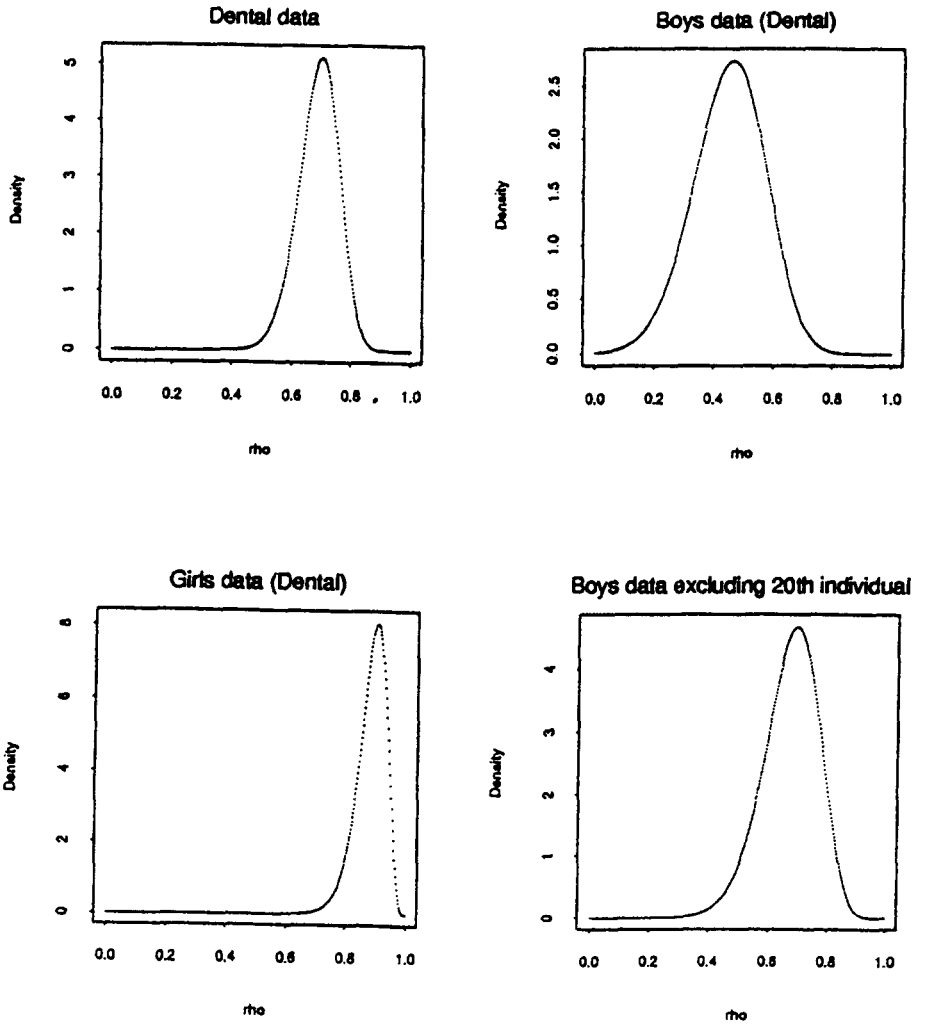


Fig. 3. Posteriors of rho.

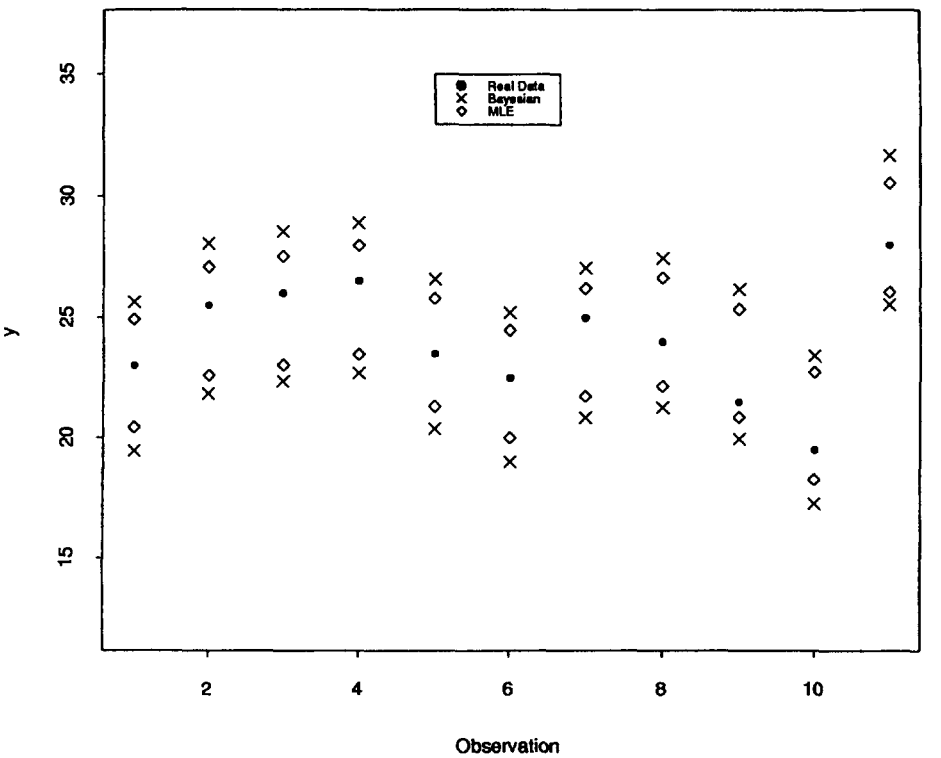


Fig. 4. Comparison of predictive intervals for y (Dental_girl data).

symmetric. This means that the corresponding approximations for the posterior distributions of τ should be quite adequate.

We begin by assuming that the girls and boys are from two different populations. The design matrix A is then a 1×11 vector for the girls and a 1×15 vector for the boys, both consisting of all 1's. In case when the individual $\#20$ is not excluded, the design matrix A is a 1×16 vector of 1's for the boys. When a common covariance structure is assumed for both populations, the design matrix A consists of 11 columns of (1, 0) followed by 15 columns of (0, 1) when the individual $\#20$ is excluded. In the situation in which the individual $\#20$ is included, then there are 16 columns of (0, 1) instead.

It is noted that when two distinct covariance structures are assumed for boys and girls, the predictions are performed separately and then the results are combined. Although the sample sizes will be smaller when compared with the common covariance structure case, the prediction performance can be better if the two convariance matrices are quite different.

The comparison of predictive performance for conditional predictions of $V^{(2)}$ given $V^{(1)}$ and Y is summarized in Table 6. The criteria used in the table are MSD, MAD and MARD. There are four different situations for this data set. Dental 1 is the dental data with girls and boys having identical covariance structure;

Table 6
Comparison of conditional predictions for four data sets

		Bayesian	ML
Dental 1	MSD	3.0135	3.3575
	MARD	0.0499	0.0527
	MAD	1.0296	1.0741
Dental 2	MSD	1.8594	1.9583
	MARD	0.0395	0.0414
	MAD	1.2899	1.3596
Dental 3	MSD	2.1636	2.5070
	MARD	0.0432	0.0451
	MAD	1.1475	1.1969
Dental 4	MSD	1.2585	1.3683
	MARD	0.0358	0.0376
	MAD	0.9435	0.9929
Ramus data	MSD	0.4996	0.5092
	MARD	0.0106	0.0107
	MAD	0.5562	0.5619
Mice data (6th)	MSD	0.0036	0.0037
	MARD	0.0606	0.0611
	MAD	0.0464	0.0465
Mice data (7th)	MSD	0.0019	0.0018
	MARD	0.0040	0.0039
	MAD	0.0362	0.0357
Glucose data (7th)	MSD	0.1047	0.1080
	MARD	0.0785	0.0803
	MAD	0.2792	0.2843
Glucose data (8th)	MSD	0.0604	0.0590
	MARD	0.0498	0.0506
	MAD	0.1924	0.1955

Note: Dental 1 is the dental data with girls and boys having identical covariance structure; Dental 2 is the same data with the individual #20 excluded; Dental 3 and Dental 4 are similar to Dental 1 and Dental 2, but with girls and boys having distinct covariance structures.

Dental 2 is the same data with the individual #20 excluded; Dental 3 and Dental 4 are similar to Dental 1 and Dental 2, but with girls and boys having distinct covariance structures. It is clear that the best situation occurs when girls and boys are assumed to have two distinct covariance structures and with the individual #20 excluded. Also, the proposed Bayesian methods are slightly better than the corresponding MLE results. As for extended predictions, the comparison is summarized in Table 7 and is expressed in terms of MAD. This table shows a similar pattern. The prediction results are best when girls and boys are assumed to have two distinct

Table 7
Comparison of extended predictions (MAD) for four data sets

	Bayesian	ML
Dental 1	1.2733	1.2903
Dental 2	1.0890	1.1059
Dental 3	1.0929	1.0989
Dental 4	0.9969	1.0117
Ramus data	0.5612	0.5655
Mice data (6th)	0.0692	0.0692
Mice data (7th)	0.0364	0.0347
Glucose data (7th)	0.2933	0.3002
Glucose data (8th)	0.1831	0.1882

Note: Dental 1–Dental 4 correspond to the four situations explained in Table 6.

covariance structures and with the individual #20 excluded. Also, the Bayesian approximations are slightly better than the MLE results for each of the situations considered.

With regard to the three other data sets, they will each be treated as from a single population. Hence, the design matrix A for each data set is self-evident. For the design matrix X, it is trivial for the ramus data, while for the glucose data we will follow Chi and Reinsel (1989) and Keramidas and Lee (1995). For the mice data, following Rao (1987) and Keramidas and Lee (1995), we will use the most recent past three observations from each mouse, i.e., $p = 3$, for both conditional and extended predictions. For conditional predictions of $Y^{(2)}$ given $Y^{(1)}$ and Y, they are done for the 6th and 7th observations for the mice data, the 7th and 8th observations for the glucose data and the 4th observations for the ramus data. The conditional prediction results are summarized in Table 6. It is clear that the proposed Bayesian methods are slightly better than the ML method for the ramus data, mice data (6th), glucose data (7th) and glucose data (8th), and are slightly worse for mice data (7th). As for extended predictions, the comparison is summarized in Table 7 and is expressed in terms of MAD. The table shows a similar pattern, i.e., the proposed Bayesian methods are slightly better than the ML method for the ramus data, mice data (6th), glucose data (7th) and glucose data (8th), and are slightly worse for mice data (7th). This perhaps means that some further modeling efforts are needed for the mice data.

Thus, it is clear that the Bayesian results developed in this paper are reasonably useful for real data. All the conditional and extended predictions are somewhat comparable to those produced by the ML method. The adequacy of the approximation for the conditional predictive density of $Y^{(2)}$ given $Y^{(1)}$ and Y is illustrated in Fig. 5 when the first column vector is treated as Y in the dental data set for girls alone.

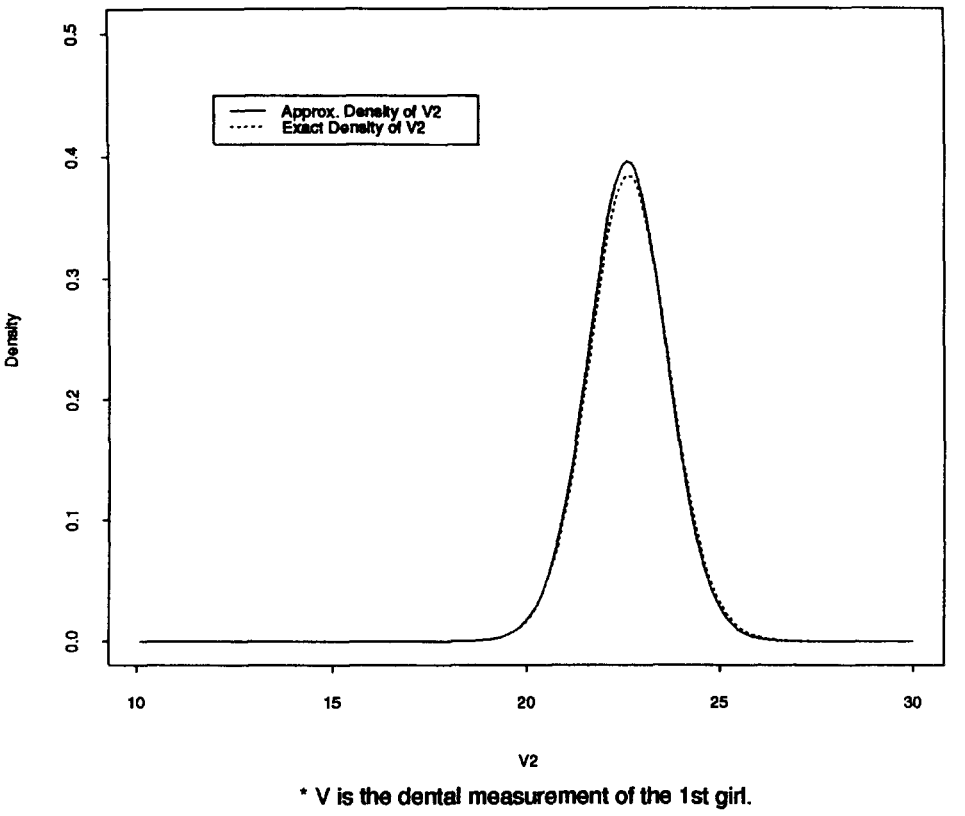


Fig. 5. Comparison of exact and approximate density of V2 given V1 and Y.

Finally, the computation involved is relatively simple and is conducted in the S-plus environment.

5. Concluding remarks

The Bayesian method presented in this paper provides an alternative way of dealing with the growth curve model when the serial covariance structure holds. The serial covariance structure is conceivably one of the most important dependence structures for this model. Although the results obtained so far are approximate in nature, they are at least comparable to those obtained by the ML method.

It is noted that the method presented in this paper provides an alternative way of constructing reliable regions for the parameters and the future values. Furthermore, the computations involved are relatively easy and should present no difficulty. It is therefore fair to say that the proposed method should be quite useful for practitioners in dealing with growth curve data.

Appendix

Theorem A.1. *Let X be distributed as χ^2_{mv}. If $\underset{m\times r}{\boldsymbol{Y}}|X = x \sim N(\boldsymbol{\mu}, b\boldsymbol{\Sigma}\otimes(x\boldsymbol{A})^{-1})$ where $\boldsymbol{\Sigma}$ is $m\times m$ p.d., $\boldsymbol{A}$ is $r\times r$ p.d., $b>0$, then the distribution of $\boldsymbol{Y}$ is given by*

$$f(\boldsymbol{Y}) = \boldsymbol{K}(m, v, r)|\boldsymbol{A}|^{m/2}b^{mv/2}|\boldsymbol{\Sigma}|^{-r/2}[b + \operatorname{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{\mu})']^{-m(v+r)/2}, \quad \text{(A.1)}$$

where

$$K(m, v, r) = \frac{\Gamma\left[\dfrac{m(v+r)}{2}\right]}{\Gamma\left[\dfrac{mv}{2}\right]}\pi^{-mr/2}. \quad \text{(A.2)}$$

Proof. The joint distribution of $\boldsymbol{Y}$ and X is

$$f(\boldsymbol{Y}, X) = \frac{(2\pi)^{-mr/2}}{\Gamma(mv/2)2^{mv/2}}b^{-mr/2}|\boldsymbol{\Sigma}|^{-r/2}|\boldsymbol{A}|^{m/2}X^{m(v+r)/2-1} \times \exp\left\{-\frac{X}{2}\left[1+\frac{1}{b}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{\mu})'\right]\right\}. \quad \text{(A.3)}$$

Next, integrating out X, we can obtain Eq. (A.1).

The density of $\boldsymbol{Y}$ as given in (A.1) is called the *trace T* distribution and will be denoted as

$$\boldsymbol{Y} \sim \operatorname{Tr}(\boldsymbol{\mu}, \boldsymbol{A}, b, \boldsymbol{\Sigma}^{-1}, m(v+r)). \quad \text{(A.4)}$$

It is noted that the distribution as given in (A.4) is one type of matrix generalization of the Students t-distribution and involves the trace of a matrix. When m or r is 1, then it will be reduced to the well-known multivariate T distribution. The first two moments of this distribution are given in the following theorem.

Theorem A.2. *If $\boldsymbol{Y} \sim \operatorname{Tr}(\boldsymbol{\mu}, \boldsymbol{A}, b, \boldsymbol{\Sigma}^{-1}, m(v+r))$, then*

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} \quad \text{(A.5a)}$$

and

$$\operatorname{Cov}(\boldsymbol{Y}) = \frac{b}{mv-2}\boldsymbol{\Sigma}\otimes\boldsymbol{A}^{-1}. \quad \text{(A.5b)}$$

Proof. The proof is completed by using the fact that

$$E(\boldsymbol{Y}) = \boldsymbol{E}_X[E(\boldsymbol{Y}|x)] \text{ and } \operatorname{Var}(\boldsymbol{Y}) = \operatorname{Var}_X[E(\boldsymbol{Y}|x)] + E_X[\operatorname{Var}(\boldsymbol{Y}|x)]$$

and $X \sim \chi^2_{mv}$ implying $E(X^{-1}) = 1/(mv-2)$. □

Theorem A.3. *If* $\boldsymbol{Y} \sim \mathrm{Tr}(\boldsymbol{\mu}, \boldsymbol{A}, b, \boldsymbol{\Sigma}^{-1}, m(v+r))$, *then*

$$U = \frac{b}{b + \mathrm{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{\mu})'} \sim \mathrm{Beta}(v_1, v_2), \tag{A.6}$$

where $v_1 = \frac{1}{2}mv$ *and* $v_2 = \frac{1}{2}mr$.

Proof. The mhth moment of U is

$$EU^{mh} = \frac{\Gamma(v_1+v_2)\Gamma(v_1+mh)}{\Gamma(v_1+v_2+mh)\Gamma(v_1)}$$

which is the mhth moment of $\mathrm{Beta}(v_1, v_2)$. Since U is a bounded random variable, its distribution is uniquely determined by its moments.

Theorem A.4. *Let*

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}^{(1)} \\ \boldsymbol{Y}^{(2)} \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}, \qquad \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix}$$

where $\boldsymbol{Y}^{(i)}, \boldsymbol{\mu}^{(i)}$ *are* $m_i \times r$; $\boldsymbol{\Lambda}_{ij}$ *is* $m_i \times m_j$; $m_1 + m_2 = m$. *If* $\boldsymbol{Y} \sim \mathrm{Tr}(\boldsymbol{\mu}, \boldsymbol{A}, b, \boldsymbol{\Sigma}^{-1}, m(v+r))$, *then the marginal distribution of* $\boldsymbol{Y}^{(2)}$ is

$$\boldsymbol{Y}^{(2)} \sim \mathrm{Tr}(\boldsymbol{\mu}^{(2)}, \boldsymbol{A}, b, \boldsymbol{\Lambda}_{22.1}, m(v+r) - m_1 r) \tag{A.7}$$

and the conditional distribution of $\boldsymbol{Y}^{(1)}|\boldsymbol{Y}^{(2)}$ *is*

$$\boldsymbol{Y}^{(1)}|\boldsymbol{Y}^{(2)} \sim \mathrm{Tr}(\boldsymbol{\mu}_{1.2}, \boldsymbol{A}, c, \boldsymbol{\Lambda}_{11}, m(v+r)), \tag{A.8}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{1.2} &= \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{Y}^{(2)} - \boldsymbol{\mu}^{(2)}), \\ c &= b + \mathrm{tr}\boldsymbol{\Lambda}_{22.1}(\boldsymbol{Y}^{(2)} - \boldsymbol{\mu}^{(2)})\boldsymbol{A}(\boldsymbol{Y}^{(2)} - \boldsymbol{\mu}^{(2)})', \\ \boldsymbol{\Lambda}_{22.1} &= \boldsymbol{\Lambda}_{22} - \boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}. \end{aligned} \tag{A.9}$$

Proof. It can be shown easily and hence is omitted.

Lemma A.1. *For the generalized growth curve model,*

$$\begin{aligned} &\mathrm{tr}\boldsymbol{C}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\tau}\boldsymbol{A})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\tau}\boldsymbol{A})' \\ &\quad = \mathrm{tr}(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})[(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1} + (\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})\boldsymbol{A}\boldsymbol{A}'(\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})'] \\ &\qquad + \mathrm{tr}(\boldsymbol{Z}'\boldsymbol{C}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}\boldsymbol{Y}'\boldsymbol{Z}, \end{aligned} \tag{A.10}$$

where

$$\boldsymbol{S} = \boldsymbol{Y}[\boldsymbol{I} - \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1}\boldsymbol{A}]\boldsymbol{Y}',$$

$$\hat{\boldsymbol{\tau}} = (\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}^{-1}\boldsymbol{Y}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1}, \tag{A.11}$$

and $\boldsymbol{Z}$ *is* $p \times (p - m)$ *and of rank* $p - m$ *such that* $\boldsymbol{X}'\boldsymbol{Z} = \boldsymbol{0}$.

It is noted that this is essentially (3.6) of Lee (1988c).

References

Chi, E.M., Reinsel, G.C., 1989. Models for longitudinal data with random effects and AR(1) errors. J. Amer. Statist. Assoc. 84, 452–459.

Elston, R.C., Grizzle, J.E., 1962. Estimation of time response curves and their confidence band. Biometrics 18, 148–159.

Fearn, T., 1975. A Bayesian approach to growth curves. Biometrika 62, 89–100.

Geisser, S., 1970. Bayesian analysis of growth curves. Sankhya, Ser. A 32, 53–64.

Geisser, S., 1980. Growth curve analysis. In: Krishnaiah, P.R., (Ed.), Handbook of Statistics, vol. 1. North-Holland, Amsterdam, pp. 89–115.

Geisser, S., 1981. Sample reuse procedures for prediction of the unobserved portion of a partially observed vector. Biometrika 68, 243–250.

Grizzle, J.E., Allen, D.M., 1969. Analysis of growth and dose response curves. Biometrics 25, 357–381.

Jennrich, R.I., Schluchter, M.D., 1986. Unbalanced repeated measures models with structured covariance matrices. Biometrics 42, 805–820.

Keramidas, E.M., Lee, J.C., 1990. Forecasting technological substitutions with concurrent short time series. J. Amer. Statist. Assoc. 85, 625–632.

Keramidas, E.M., Lee, J.C., 1995. Selection of a covariance structure for growth curves. Biometrical J. 37, 783–797.

Khatri, C.G., 1966. A note on a MANOVA model applied to problems in growth curves. Ann. Inst. Statist. Math. 18, 75–86.

Khatri, C.G., 1973. Testing some covariance structures under a growth curve model. J. Multivariate Anal. 3, 102–116.

Laitinen, K., 1978. Why is demand homogeneity so often rejected? Economics Lett. 1, 187–191.

Lee, J.C., 1982. Classification of growth curves. In: Krishnaiah, P.R., Kanal, L.W. (Eds.), Handbook of Statistics, vol. 2. North-Holland, Amsterdam, pp. 121–137.

Lee, J.C., 1988a. Growth curve model and technological forecasting. In: Matusita, K., (Ed.), Statistical Theory and Data Analysis, vol. 2. North-Holland, Amsterdam, pp. 369–378.

Lee, J.C., 1988b. Nested rotterdam model: applications to marketing research with special reference to telecommunications demand. Internat. J. Forecasting 4, 193–206.

Lee, J.C., 1988c. Prediction and estimation of growth curves with special covariance structures. J. Amer. Statist. Assoc. 83, 432–440.

Lee, J.C., 1991. Test and model selection for the general growth curve model. Biometrics 47, 147–159.

Lee, J.C., Geisser, S., 1972. Growth curve prediction. Sankhya Ser. A 34, 393–412.

Lee, J.C., Geisser, S., 1975. Applications of growth curve prediction. Sankhya Ser. A 37, 239–256.

Lee, J.C., Geisser, S., 1996. On the prediction of growth curves In: Lee, J.C., Zellner, A., Johnson, W.O. (Eds.), Modelling and Prediction Honoring Seymour Geisser. Springer, Berlin, 71–103.

Ljung, G.M., Box, G.E.P., 1980. Analysis of variance with autocorrelated observations. Scand. J. Statist. 7, 172–180.

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

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: Krishnaiah, P.R., (Ed.), Multivariate Analysis, vol. 1. Academic Press, New York, pp. 87–103.

Rao, C.R., 1967. Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In: LeCam, L.M., Neyman, J. (Eds.), Proc. 5th Berkeley Symp. Mathematical Statistics and Probability, vol. 1. University of California Press, Berkeley: pp. 355–372.

Rao, C.R., 1977. Prediction of future observations with special reference to linear models. In: Krishnaiah, P.R., (Ed.), Multivariate Analysis, vol. 4. Academic Press, New York, pp. 193–208.

Rao, C.R., 1984. Prediction of future observations in polynomial growth curve models. In: Proc. Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions, Indian Statistical Institute, Calcutta, pp. 512–520.

Rao, C.R., 1987. Prediction of future observations in growth curve models. Statist. Sci. 2, 434–471.

Williams, J.S., Izenman, A.J., 1981. A class of linear spectral models and analysis for the study of longitudinal data. Technical Report, Dept. of Statistics, Colorado State University.

Zellner, A., Tiao, G.C., 1964. Bayesian analysis of the regression model with autocorrelated errors. J. Amer. Statist. Assoc. 59, 763–778.

Zerbe, G.O., 1979. Randomization analysis of the completely randomized design extended to growth and response curves. J. Amer. Statist. Assoc. 74, 215–221.