 Vol 27: 703–713, 2000

Bayesian Analysis of a Growth Curve Model with a General Autoregressive Covariance Structure

J. C. LEE and C. H. CHANG

National Chiao Tung University

ABSTRACT. In this paper we consider from maximum likelihood and Bayesian points of view the generalized growth curve model when the covariance matrix has a Toeplitz structure. This covariance is a generalization of the AR(1) dependence structure. Inferences on the parameters as well as the future values are included. The results are illustrated with several real data sets.

Key words: banded covariance, Bayesian, exact and approximate distributions, maximum likelihood estimates, MCMC, non-informative prior, prediction, Toeplitz structure

1. Introduction

We consider a generalized multivariate analysis of variance model useful especially for many types of growth curve problems. The generalized growth curve model, proposed by Potthoff & Roy (1964), is defined as

$$\underset{p\times N}{\mathbf{Y}} = \underset{p\times m}{\mathbf{X}}\ \underset{m\times r}{\tau}\ \underset{r\times N}{\mathbf{A}} + \underset{p\times N}{\varepsilon} \tag{1}$$

where τ is unknown and $\mathbf{X}$ and $\mathbf{A}$ are known design matrices of ranks $m < p$ and $r < N$, respectively. The columns of ε are independent p-variate normal with mean vector $\mathbf{0}$ and common covariance matrix Σ. In general, p is the number of time (or spatial) points observed on each of 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 polynomial for the growth function and the distinct grouping out of N independent vector observations. For some reviews, see Geisser (1980), Von Rosen (1991) and Kshirsagar & Smith (1995). For related literature on repeated measurements, see Lindsey (1993).

When Σ is arbitrary and positive definite, the most general case, the estimation for $\boldsymbol{\tau}$ and Σ as well as the prediction problem to be addressed in this paper, the reader is referred to Khatri (1966), Geisser (1970) and Lee & Geisser (1972). For the situation in which some parsimonious covariance structure is more appropriate for the data at hand, then the arbitrary and positive definite Σ will not be suitable for the data set. The covariance structure considered in this paper is quite reasonable for a lot of growth curve data.

The general autoregressive, or banded, covariance structure, is defined as

$$\Sigma = \sigma^2\mathbf{C} \tag{2}$$

where $\mathbf{C} = (c_{ij})$, $c_{ij} = \rho_{|i-j|}$, $i \neq j$, $c_{ii} = 1$, for $i, j = 1, \ldots, p$, $\sigma^2 > 0$ and $\rho_{|i-j|}$ are unknown and $-1 < \rho_{|i-j|} < 1$ subject to $\mathbf{C}$ being positive definite. This covariance structure is considered by Lee & Geisser (1975) offering a rough solution, and Jennrich & Schluchter (1986) offering numerical search solution in a general setting. In this paper, we let $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_{p-1})$. In addition to the AR(1) dependence, or serial covariance structure, it is perhaps one of the more important covariance structures for the generalized growth curve model. The purpose of this paper is to consider parameter estimation and prediction of future values from maximum likelihood (ML) and Bayesian points of view. In the Bayesian treatment of the model, we will consider both simple approximation which is conditional in nature, as done in Lee & Hsu (1997) for the AR(1) dependence, and Markov chain Monte Carlo (MCMC) methods. The Bayesian approach is emphasized because it will be useful when the sample size is relatively small. Indeed, several published data sets are relatively small in their sizes. We will compare our results obtained from both ML and Bayesian approaches via several data sets.

In addition to the inferences of 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)–(2). Let $\mathbf{V}$ be $p \times K$ future observations drawn from the generalized growth curve model; that is, the future observations are such that given the parameters τ and Σ,

$$\mathbf{V} = \mathbf{X}\tau\mathbf{F} + \varepsilon^* \tag{3}$$

where $\mathbf{F}$ is a known $r \times K$ matrix, and the columns of ε^* are independent and p-variate normal with mean vector $\mathbf{0}$ and common covariance matrix Σ. Geisser (1970, 1980) and Lee (1982) considered prediction of $\mathbf{V}$, given $\mathbf{Y}$ as the sample from a Bayesian viewpoint. Lee & Geisser (1972, 1975), Fearn (1975), Rao (1987), and Lee (1988) considered the problem of predicting $\mathbf{V}^{(2)}$, given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$, if $\mathbf{V}$ is partitioned as $\mathbf{V} = (\mathbf{V}^{(1)\prime}, \mathbf{V}^{(2)\prime})'$, 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 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.

In section 2, we consider parameter estimation and prediction of future values based on the ML method. In section 3, Bayesian estimation of the parameters and prediction of the future values are considered via Markov chain Monte Carlo (MCMC) methods and simple approximations. The results developed in the paper are illustrated in section 4 with real data. Finally, some concluding remarks are given in section 5.

2. Estimation and prediction based on the ML method

2.1. Parameter estimation with $\mathbf{Y}$ *and* $\mathbf{V}^{(1)}$ *as the sample*

In this section we will consider the situation in which both $\mathbf{Y}$ and the partially observed vector, $\mathbf{V}^{(1)}$, are used as the sample in the estimation of parameters and for the predictive inference of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$. In case only $\mathbf{Y}$ is used as the sample, the estimation results can be obtained from theorem 1 by setting $p_1 = 0$ and deleting the terms involving $\mathbf{V}^{(1)}$.

Theorem 1

For the growth curve model, when covariance matrix Σ satisfies the structure given by (2) *and $r = 1$, the MLEs of τ and σ^2, denoted by $\hat{\tau}$ and $\hat{\sigma}^2$, respectively, are*

$$\hat{\tau} = \mathbf{Q}^{-1}(\mathbf{Q}_1\hat{\tau}_1 + \mathbf{Q}_2\hat{\tau}_2),$$

$$\hat{\sigma}^2 = [(\hat{\tau}_1 - \hat{\tau}_2)'\mathbf{Q}_1\mathbf{Q}^{-1}\mathbf{Q}_2(\hat{\tau}_1 - \hat{\tau}_2) + \mathrm{tr}(\mathbf{X}'\hat{\mathbf{C}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{C}}^{-1}\mathbf{S}\hat{\mathbf{C}}^{-1}\mathbf{X}$$
$$+ \mathrm{tr}(\mathbf{Z}'\hat{\mathbf{C}}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}\mathbf{Y}'\mathbf{Z} + \mathrm{tr}(\mathbf{X}^{(1)'}\hat{\mathbf{C}}_{11}^{-1}\mathbf{X}^{(1)})^{-1}\mathbf{X}^{(1)'}\hat{\mathbf{C}}_{11}^{-1}\mathbf{S}_1\hat{\mathbf{C}}_{11}^{-1}\mathbf{X}^{(1)}$$
$$+ \mathrm{tr}(\mathbf{Z}_1'\hat{\mathbf{C}}_{11}\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{V}^{(1)}\mathbf{V}^{(1)'}\mathbf{Z}_1]/(pN + p_1K), \tag{4}$$

where

$$Q_1 = \mathbf{A}\mathbf{A}'(\mathbf{X}'\hat{\mathbf{C}}^{-1}\mathbf{X}), \quad \mathbf{Q}_2 = \mathbf{F}\mathbf{F}'(\mathbf{X}^{(1)}\hat{\mathbf{C}}_{11}^{-1}\mathbf{X}^{(1)}),$$
$$\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2,$$
$$\hat{\tau}_1 = (\mathbf{X}'\hat{\mathbf{C}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{C}}^{-1}\mathbf{Y}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1},$$
$$\hat{\tau}_2 = (\mathbf{X}^{(1)'}\hat{\mathbf{C}}_{11}^{-1}\mathbf{X}^{(1)})^{-1}\mathbf{X}^{(1)'}\hat{\mathbf{C}}_{11}^{-1}\mathbf{V}^{(1)}\mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1},$$

$\mathbf{Z}_1$ *is a known* $p_1 \times (p_1 - m)$ *matrix with rank* $p_1 - m$ *such that* $\mathbf{X}^{(1)'}\mathbf{Z}_1 = \mathbf{0}$,

$$\mathbf{S}_1 = \mathbf{V}^{(1)}[\mathbf{I} - \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}\mathbf{F}]\mathbf{V}^{(1)'},$$

$\hat{\mathbf{C}} = (\hat{\mathbf{C}}_{ab})$, $\hat{\mathbf{C}}_{ab}$ is of dimension $p_a \times p_b$, $p_a + p_b = p$,

$\hat{\mathbf{C}} = (\hat{c}_{ij})$, $\hat{c}_{ij} = \hat{\rho}_{|i-j|}$, $i \neq j$, *and* $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_{p-1})$, *the MLE of* ρ, *is obtained by maximizing* $L(\rho|\mathbf{Y}, \mathbf{V}^{(1)})$,

$$L(\rho|\mathbf{Y}, \mathbf{V}^{(1)}) \propto b_{\mathbf{V}^{(1)}}^{-[(pN+p_1K)/2]}|\mathbf{C}|^{-(N/2)}|\mathbf{C}_{11}|^{-(K/2)}, \tag{5}$$

subject to $\mathbf{C}$ *being positive definite and*

$$b_{\mathbf{V}^{(1)}} = (\hat{\tau}_1 - \hat{\tau}_2)'\mathbf{Q}_1\mathbf{Q}^{-1}\mathbf{Q}_2(\hat{\tau}_1 - \hat{\tau}_2) + \mathrm{tr}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{S}\mathbf{C}^{-1}\mathbf{X}$$
$$+ \mathrm{tr}(\mathbf{Z}'\mathbf{C}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}\mathbf{Y}'\mathbf{Z} + \mathrm{tr}(\mathbf{Z}_1'\mathbf{C}_{11}\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{V}^{(1)}\mathbf{V}^{(1)'}\mathbf{Z}_1.$$

2.2. Prediction of $\mathbf{V}^{(2)}$ *given* $\mathbf{V}^{(1)}$ *and* $\mathbf{Y}$

The approximate mean, denoted by $\hat{\mathbf{V}}^{(2)}$, of the distribution of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$ is

$$\hat{\mathbf{V}}^{(2)} = \mathbf{X}^{(2)}\hat{\tau}\mathbf{F} + \hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}(\mathbf{V}^{(1)} - \mathbf{X}^{(1)}\hat{\tau}\mathbf{F}), \tag{6}$$

where $\mathbf{X} = (\mathbf{X}^{(1)'}, \mathbf{X}^{(2)'})'$, $\hat{\Sigma} = \hat{\sigma}^2\hat{\mathbf{C}} = (\hat{\Sigma}_{ij})$, $\hat{\tau}$ and $\hat{\sigma}^2$ are given in (4), $\mathbf{X}^{(i)}$ is $p_i \times m$, and $\hat{\Sigma}_{ij}$ is of dimension $p_i \times p_j$, $p_1 + p_2 = p$.

3. Bayesian inferences of the model

3.1. Parameter estimation

The likelihood of τ, σ^2 and ρ is

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

For the prior of τ, σ^2 and ρ, we will use the following non-informative prior

$$g(\tau, \sigma^2, \rho) \propto \frac{1}{\sigma^2}. \tag{7}$$

In (7), 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 & Tiao (1964).

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

$$P(\tau, \sigma^2, \rho|\mathbf{Y}) \propto \sigma^{-(pN+2)}|\mathbf{C}|^{-(N/2)} \times \exp\left\{-\frac{1}{2\sigma^2}\operatorname{tr}\mathbf{C}^{-1}\left[\mathbf{Y}-(\mathbf{X},\mathbf{Z})\begin{pmatrix}\tau\\ \mathbf{0}\end{pmatrix}\mathbf{A}\right]\left[\mathbf{Y}-(\mathbf{X},\mathbf{Z})\begin{pmatrix}\tau\\ \mathbf{0}\end{pmatrix}\mathbf{A}\right]'\right\}. \tag{8}$$

Integration of (8) w.r.t. τ, using Lee (1988, (3.6)) and the application of some algebraic identities yield

$$P(\tau, \rho|\mathbf{Y}) \propto |\mathbf{C}|^{-(N/2)} S_1^{-(pN/2)}, \tag{9}$$

where

$$S_1 = \operatorname{tr}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}(\tau-\hat{\tau}_0)\mathbf{A}\mathbf{A}'(\tau-\hat{\tau}_0)' + b_0,$$

$$b_0 = \operatorname{tr}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{S}\mathbf{C}^{-1}\mathbf{X} + \operatorname{tr}(\mathbf{Z}'\mathbf{C}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}\mathbf{Y}'\mathbf{Z}$$

and $\hat{\tau}_0$ is same as $\hat{\tau}_1$ in (4) with $\hat{\mathbf{C}}$ replaced by $\mathbf{C}$.
From (9) we see that conditional on ρ,

$$P(\tau|\mathbf{Y}, \rho) = \operatorname{tr}(\hat{\tau}_0, \mathbf{A}\mathbf{A}', b_0, \mathbf{X}'\mathbf{C}^{-1}\mathbf{X}, pN). \tag{10}$$

Here $\tau_{m\times r}$ is distributed as $\operatorname{tr}(\mu, \mathbf{B}, b, \Sigma^{-1}, m(r+\nu))$ if its pdf is

$$f(\tau) = K(m, \nu, r)|\mathbf{B}|^{m/2} b^{m\nu/2}|\Sigma|^{-(r/2)}[b + \operatorname{tr}\Sigma^{-1}(\tau-\mu)\mathbf{B}(\tau-\mu)']^{-[m(r+\nu)/2]}, \tag{11}$$

where

$$K(m, \nu, r) = \Gamma(m(r+\nu)/2)\Gamma^{-1}(m\nu/2)\pi^{-m\nu/2}.$$

The density of τ as given in (11) is called the trace T distribution by Lee & Hsu (1997) in which some properties of the distribution are discussed.

Moreover, integrating out τ in (9) we have the following posterior density of ρ:

$$P(\rho|\mathbf{Y}) \propto b^{-(pN-mr/2)}|\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}|^{-(r/2)}|\mathbf{C}|^{-(N/2)}. \tag{12}$$

Since

$$P(\tau|\mathbf{Y}) = \int P(\tau, \rho|\mathbf{Y})d\rho = \int P(\tau|\mathbf{Y}, \rho)P(\rho|\mathbf{Y})d\rho, \tag{13}$$

the posterior density of τ can be approximated by

$$P(\tau|\mathbf{Y}) \doteq P(\tau|\mathbf{Y}, \hat{\rho}), \tag{14}$$

where ρ maximizes the posterior density of ρ, as given in (12). This approximation will be reasonable if $P(\rho|\mathbf{Y})$ is concentrated and nearly symmetric, as pointed out by Ljung & Box (1980). Thus, approximately, the posterior distribution of τ is a *trace T* distribution. Hence, a $1-\alpha$ posterior region for τ can be obtained from the following inequality:

$$\hat{b}_0^{-1}\operatorname{tr}(\mathbf{X}'\hat{\mathbf{C}}^{-1}\mathbf{X})(\tau-\hat{\tau}_0^*)\mathbf{A}\mathbf{A}'(\tau-\hat{\tau}_0^*)' \leqslant \frac{mr}{pN-mr}F_\alpha(mr, pN-mr), \tag{15}$$

where $\hat{\tau}_0^*$, $\hat{b}_0$, $\hat{\mathbf{C}}$ are the $\hat{\tau}_0$, b_0 and $\mathbf{C}$ evaluated at $\rho = \hat{\rho}$ and $F_\alpha(\nu_1, \nu_2)$ is the upper 100α per cent point of the F distribution. In (13), the integration can be carried out numerically and will be considered as the "exact" posterior density. A better approximation than (14) would be the Rao–Blackwellization approximation

$$P(\tau|\mathbf{Y}) \doteq \frac{1}{L}\sum_{i=1}^{L} P(\tau|\mathbf{Y}, \rho^{(i)}), \tag{16}$$

where $\rho^{(i)}$ is the ith draw from $P(\rho|\mathbf{Y})$.

Integration w.r.t. τ and ρ in (2) and using arguments similar to (14), we obtain the approximate posterior distribution of σ^2 as

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

where $\mathrm{IG}(\nu_1, \nu_2)$ is the inverse gamma distribution with parameters ν_1 and ν_2.

3.2. *Predictive inference of future values* V

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

$$f(\mathbf{V}|\tau, \sigma^2, \rho) \propto \sigma^{-pK}|\mathbf{C}|^{-(K/2)} \exp\left[-\frac{1}{2\sigma^2}\operatorname{tr}\mathbf{C}^{-1}(\mathbf{V} - \mathbf{X}\tau\mathbf{F})(\mathbf{V} - \mathbf{X}\tau\mathbf{F})'\right]. \tag{17}$$

Upon combining with the posterior density of τ, σ^2 and ρ, as given in (8) and integrating w.r.t. σ^2 and τ, we have

$$P(\mathbf{V}, \rho|\mathbf{Y}) \propto b_1{}^{-[(p(N+K)-mr)/2]}|\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}|^{-(r/2)}|\mathbf{C}|^{-(N+K/2)}, \tag{18}$$

where $b_1 = b_0 + \operatorname{tr}\mathbf{G}(\mathbf{V} - \mathbf{X}\hat{\tau}\mathbf{F})(\mathbf{V} - \mathbf{X}\hat{\tau}\mathbf{F})'$, $\mathbf{G} = \mathbf{M}\mathbf{C}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1} + \mathbf{Z}(\mathbf{Z}'\mathbf{C}\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{M} = \mathbf{I} - \mathbf{F}'(\mathbf{A}_0\mathbf{A}_0)^{-1}\mathbf{F}$, $\mathbf{A}_0 = (\mathbf{A}, \mathbf{F})$, and b_0 is defined in (9). From (18) we see that $P(\mathbf{V}|\mathbf{Y}, \rho)$ is a trace T distribution. Hence, approximately,

$$P(\mathbf{V}|\mathbf{Y}) \doteq P(\mathbf{V}|\mathbf{Y}, \rho) = \operatorname{tr}(\mathbf{X}\hat{\tau}\mathbf{F}, \hat{b}_0, \mathbf{G}, p(N+K) - mr), \tag{19}$$

where ρ maximizes

$$P(\rho|\mathbf{Y}) \propto |\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}|^{-(r/2)}|\mathbf{C}|^{-((N+K)/2)} b^{-((pN-mr)/2)}|\mathbf{G}^{-(p/2)}, \tag{20}$$

and $\hat{b}_0$ is the b_0 defined in (9) evaluated at $\rho = \hat{\rho}$.

An approximate $1 - \alpha$ predictive region for $\mathbf{V}$ can be obtained through

$$\hat{b}_0^{-1}\operatorname{tr}\hat{\mathbf{G}}(\mathbf{V} - \mathbf{X}\hat{\tau}\mathbf{F})(\mathbf{V} - \mathbf{X}\hat{\tau}\mathbf{F})' \leqslant \frac{pK}{pN - mr} F_\alpha(pK, pN - mr). \tag{21}$$

We next consider the conditional predictive density of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$. From (18) we get

$$P(\mathbf{V}^{(2)}, \rho|\mathbf{V}^{(1)}, \mathbf{Y}) \propto |\mathbf{C}|^{-((N+K)/2)}|\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}|^{-(r/2)} \times [b_2 + \operatorname{tr}(\mathbf{V}^{(2)} - \hat{\mathbf{V}}^{(2)})'\mathbf{G}_{22}(\mathbf{V}^{(2)} - \hat{\mathbf{V}}^{(2)}]^{-((p(N+K)-mr)/2)} \tag{22}$$

where $b_2 = b_0 + \operatorname{tr}(\mathbf{V}^{(1)} - \mathbf{X}^{(1)}\hat{\tau}\mathbf{F})'\mathbf{G}_{11.2}(\mathbf{V}^{(1)} - \mathbf{X}^{(1)}\hat{\tau}\mathbf{F})$,

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

$$\mathbf{G} = (G_{ij}), \quad i, j = 1, 2, \quad \mathbf{G}_{11.2} = \mathbf{G}_{11} - \mathbf{G}_{12}\mathbf{G}_{22}^{-1}\mathbf{G}_{21}, \quad p_1 + p_2 = p. \tag{23}$$

It is clear that conditional on ρ,

$$P(\mathbf{V}^{(2)}|\rho, \mathbf{V}^{(1)}, \mathbf{Y}) = \operatorname{tr}(\hat{\mathbf{V}}^{(2)}, \mathbf{I}, b_2, \mathbf{G}_{22}, p(N+K) - mr). \tag{24}$$

Meanwhile, integrating out $\mathbf{V}^{(2)}$ in (22), we have

$$P(\rho|\mathbf{V}^{(1)}, \mathbf{Y}) \propto b_2^{-((pN+p_1K-mr)/2)}|\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}|^{-(r/2)}|\mathbf{C}|^{-(N+K/2)}|\mathbf{G}_{22}|^{-(p_2/2)}. \tag{25}$$

As in (14), we have

$$P(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y}) \doteq P(\mathbf{V}^{(2)}|\hat{\rho}, \mathbf{V}^{(1)}, \mathbf{Y}), \tag{26}$$

where ρ maximizes (25) subject to $\mathbf{C}$ being positive definite and when $P(\rho|\mathbf{V}^{(1)}, \mathbf{Y})$ is symmetric and concentrated.

Thus, we obtain the following approximation,

$$P(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y}) \doteq \text{tr}(\hat{\mathbf{V}}^{(2)*}, \mathbf{I}, \hat{b}_2, \hat{\mathbf{G}}_{22}, p(N+K) - mr), \tag{27}$$

where $\hat{\mathbf{V}}^{(2)*}$, $\hat{b}_2$, and $\hat{\mathbf{G}}_{22}$ are the $\hat{\mathbf{V}}^{(2)}$, b_2, and $\mathbf{G}_{22}$, defined in (23), evaluated at $\rho = \hat{\rho}$, and $\hat{\rho}$ maximizes $P(\rho|\mathbf{V}^{(1)}, \mathbf{Y})$ subject to C being positive definite.

An approximate $1 - \alpha$ predictive region for $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$ can be obtained through

$$\hat{b}_2^{-1} \,\text{tr}\, \hat{\mathbf{G}}_{22}(\mathbf{V}^{(2)} - \hat{\mathbf{V}}^{(2)*})(\mathbf{V}^{(2)} - \hat{\mathbf{V}}^{(2)*})' \leqslant \frac{p_2 K}{pN + p_1 K - mr} F_\alpha(p_2 K, pN + p_1 K - mr). \tag{28}$$

On the other hand, we can integrate $P(\mathbf{V}^{(2)}, \rho|\mathbf{V}^{(1)}, \mathbf{Y})$ with respect to ρ to obtain the "exact" predictive distribution of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$. We can then compare the "exact" predictive distribution with the approximation as given in (27).

3.3. *MCMC predictive inference of* $\mathbf{V}^{(2)}$ *given* $\mathbf{V}^{(1)}$ *and* $\mathbf{Y}$

For ease of presentation and some practical consideration, we will restrict our attention to the special situation in which $r = K = 1$ for the rest of the paper. We will next describe the non-trivial part in the MCMC methodology and forecasting procedure and discuss the MCMC approximation to the predictive distribution of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$. For more details on MCMC, see Metropolis *et al.* (1953), Hastings (1970), Gelfand & Smith (1990), Gelfand *et al.* (1990), Casella & George (1992), and Gilks *et al.* (1996), among others.

3.3.1. *Forecast*

Once the posterior distribution of the parameters are obtained through the MCMC sampler, we can use it to predict the future values of $\mathbf{Y}$. Suppose we are at the nth period. Let D_n denote the data set $\{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n\}$. Let θ denote the parameters. Prediction for the $(n+1)$th period follows from the predictive density

$$f(\mathbf{Y}_{n+1}|D_n) = \int f(\mathbf{Y}_{n+1}|D_n, \theta)\pi(\theta|D_n)d\theta, \tag{29}$$

where $\mathbf{Y}_{n+1}$ denotes the random future observation at period $n+1$. This density can be approximated by Monte Carlo integration using the MCMC samples

$$\hat{f}(\mathbf{Y}_{n+1}|D_n) = \frac{1}{L}\sum_{s=1}^{L} f(\mathbf{Y}_{n+1}|D_n, \theta^{(k,s)}). \tag{30}$$

The mean of this predictive distribution is computed from

$$E(\mathbf{Y}_{n+1}|D_n) = E(E(\mathbf{Y}_{n+1}|D_n, \theta)|D_n). \tag{31}$$

3.3.2. *MCMC approximation of posterior density of* $\mathbf{V}^{(2)}$ *given* $\mathbf{V}^{(1)}$ *and* $\mathbf{Y}$

We can use the MCMC sampler to obtain an approximate posterior density of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$ by

$$P(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y}) \doteq \frac{1}{L}\sum_{s=1}^{L} P(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y}, \theta^{(k,s)}), \tag{32}$$

where $\theta^{(k,s)} = (\tau^{(k,s)}, (\sigma^2)^{(k,s)}, \rho^{(k,s)})$ are the realizations of τ, σ^2, ρ for the kth iteration and sth replication, respectively, and the approximate predictor of $\mathbf{V}^{(2)}$ is computed from

$$\hat{\mathbf{V}}^{(2)} \doteq \frac{1}{L}\sum_{s=1}^{L} (\hat{\mathbf{V}}^{(2)})^{(k,s)},$$

where $(\mathbf{V}^{(2)})^{(k,s)}$ is the realization of $\mathbf{V}^{(2)}$ for the kth iteration and sth replication.

The approximation as given in (32) is expected to be better than (27) and an illustration is shown in Fig. 1. In the MCMC approximation (32), the generation of σ^2, τ, $\mathbf{V}^{(2)}$ will be self-evident. As for ρ, we generate ρ given τ, σ^2, $\mathbf{V}^{(2)}$, $\mathbf{V}^{(1)}$, $\mathbf{Y}$ using the Metropolis algorithm where

$$f(\rho|\tau, \sigma^2, \mathbf{V}, \mathbf{Y}) \propto |\mathbf{C}|^{-((N+1)/2)} \exp\left[-\frac{1}{2\sigma^2}\operatorname{tr}\mathbf{C}^{-1}(\mathbf{Y}_0 - \mathbf{X}\tau\mathbf{A}_0)(\mathbf{Y}_0 - \mathbf{X}\tau\mathbf{A}_0)'\right],$$

$-1 < \rho_i < 1$, for $i = 1, \ldots, p-1$.

To elaborate on the Metropolis algorithm in generating ρ, let us assume that the prior on ρ_i is uniform over $(-1, 1)$ for $i = 1, \ldots, p-1$. We can transform ρ_i to $\rho_i' \in (-\infty, \infty)$ by

$$\rho_i' = \log\left(\frac{1+\rho_i}{1-\rho_i}\right), \quad \text{for } i = 1, \ldots, p-1.$$

Then

$$\rho_i = \frac{\exp(\rho_i') - 1}{\exp(\rho_i') + 1}$$

and the Jacobian of the transformations from $\rho = (\rho_1, \ldots, \rho_{p-1})$ to $\rho' = (\rho_1', \ldots, \rho_{p-1}')$ is

$$J = \prod_{l=1}^{p-1} \frac{2\exp(\rho_1')}{(\exp(\rho_1') + 1)^2}.$$

Hence the conditional density of ρ' is

$$f_{\rho'}(\rho') = f_\rho(\rho(\rho'))\cdot J. \tag{33}$$

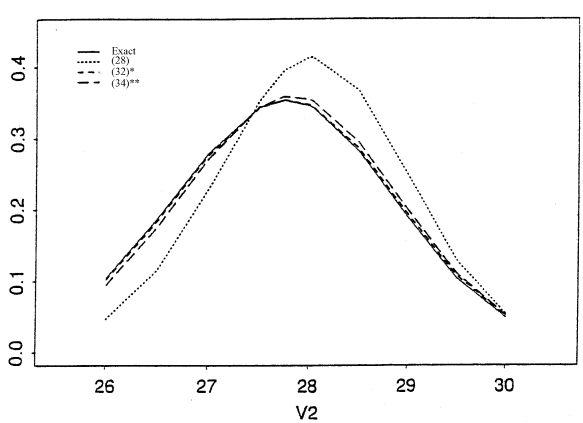

Fig. 1. Comparison of exact and approximate predictive distributions of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$. $\mathbf{V}$ is the last vectorial observation of the girl data and $\mathbf{V}^{(2)}$ is the last component of $\mathbf{V}$

Then we apply the Metropolis algorithm to the probability density function of ρ'. Here we must note that we define a transition kernel $q(\rho', \mathbf{y})$ such that $\mathbf{y} = \rho' + \Omega^{1/2}\mathbf{Z}$ with $\mathbf{Z}$ being the standard multivariate normal random variates and Ω reflecting the conditional covariance of ρ' in (33). We can estimate the conditional covariance matrix Ω of ρ' by inverting the sample information matrix at the generated τ, σ^2, ρ and $\mathbf{V}^{(2)}$.

An alternative approximation for $P(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y})$ is

$$P(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y}) \doteq \frac{1}{L}\sum_{i=1}^{L} P(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y}, \rho^{(i)}), \tag{34}$$

where $\rho^{(i)}$ is the ith draw from

$$f_\rho^*(\rho) = P(\rho|\mathbf{V}^{(1)}, \mathbf{Y}) \propto |\mathbf{C}|^{-((N+1)/2)}|\mathbf{X}'\mathbf{C}^{-1}\mathbf{X}|^{-1/2} b_1^{-((pN+p_1-mr)/2)}|\mathbf{G}_{22}|^{-(p_2/2)},$$

which can be generated via the Metropolis algorithm.

The mean of this approximate predictive distribution is computed from

$$E(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y}) = E(E(\mathbf{V}^{(2)}|\mathbf{V}^{(1)}, \mathbf{Y}, \rho)|\mathbf{V}^{(1)}, \mathbf{Y}).$$

As before, we transform ρ_i to ρ_i' for $i = 1, \ldots, p-1$, and then apply the Metropolis algorithm to generate ρ'. We can then estimate the conditional covariance matrix Ω of ρ' by inverting the sample information matrix at the generated ρ.

4. Numerical illustrations

For illustration purposes, we will apply some of the results developed in sections 2 and 3 to four biological data sets including three sets (dental data with 16 boys and 11 girls, ramus data and mice data) analysed by Rao (1987) and Lee (1988) and one set (glucose data) analysed by Chi & Reinsel (1989). For the dental data, since individual 20 is suspected to be an aberrant observation (Lee & Geisser, 1975), the prediction comparison will be considered for the situation in which this particular observation is excluded as well. In this situation, $N = 26$ for the dental data and $N = 15$ for the boy data.

It is noted that in all four data sets the covariate is time. Also, our purpose here is for illustration showing that the results presented in earlier sections can be implemented. Of course, for a data set if an AR(1) dependence is suitable, then the general autoregressive covariance structure considered in this paper should be a possible candidate as well.

Similar to Lee & Geisser (1975), Fearn (1975) and Lee (1988), we will predict the last observation of a partially observed vector, that is, $K = 1$, $p_2 = 1$, and $p_1 = p - 1$. For prediction purposes, we withhold one vector and use the rest for predicting the last component of that vector, and repeat this for each of the N observations. This gives N predicted values for the last N observed values. The mean squared deviation (MSD), the mean absolute deviation (MAD), and the mean absolute relative deviation (MARD) of the predicted values from the actual observations are used to assess the relative merits of the various predictors.

From the prediction results given in Table 1, it is clear that for the dental data set, the individual 20, which is a boy, does contribute heavily to the prediction error as seen from MSD, MAD and MARD. For example, in terms of MAD, it increases from 1.2211 to 1.3759 for the Bayesian result which means that the absolute prediction error for this individual is 5.4007. This is much bigger than the MAD of 1.2211 when the individual 20 is excluded. Also, for the glucose data, the last time point ($p = 8$) is harder to predict than the previous three time points ($p = 5, 6, 7$). Overall, the Bayesian results are somewhat comparable or slightly better than those using the ML method.

Table 1. *Comparison of conditional predictions: based on* $\mathbf{V}^{(1)}$ *and* $\mathbf{Y}$ *as the sample*

	ML			Bayes		
	MSD	MAD	MARD	MSD	MAD	MARD
Dental data, N = 27	3.2131	1.3854	0.0543	3.1317	1.3759	0.0538
Dental data, N = 26	2.5488	1.2296	0.0482	2.4893	1.2211	0.0479
Boy data, N = 16	4.0979	1.5828	0.0593	3.8780	1.5674	0.0584
Boy data, N = 15	2.8701	1.3544	0.0503	2.6648	1.3202	0.0489
Girl data	1.0570	0.8676	0.0369	1.0100	0.8568	0.0365
Ramus Bone data	0.7717	0.6268	0.0121	0.7698	0.6209	0.0120
Mice data						
$p = 4$	0.0024	0.0430	0.0458	0.0024	0.0426	0.0455
$p = 5$	0.0023	0.0390	0.0426	0.0024	0.0410	0.0466
$p = 6$	0.0027	0.0454	0.0501	0.0026	0.0441	0.0489
$p = 7$	0.0024	0.0412	0.0444	0.0023	0.0408	0.0441
Glucose data						
$p = 5$	0.0949	0.2530	0.0675	0.0946	0.2513	0.0670
$p = 6$	0.0797	0.2342	0.0638	0.0798	0.2341	0.0638
$p = 7$	0.0874	0.2467	0.0654	0.0880	0.2470	0.0657
$p = 8$	0.1189	0.2659	0.0734	0.1174	0.2643	0.0730

For illustration purposes, we use the last vectorial observation of the girl data as $\mathbf{V}$ and the rest as $\mathbf{Y}$ to obtain the approximate predictive density of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$, when $p_1 = 1$. The comparison among the exact, obtained by numerical integration, approximations (27), (32) and (34) is given in Fig. 1. It is seen that for this data set the simple approximation (27) is not adequate while (32) and (34) are almost as good as the exact. The MCMC approximations to the predictive distribution of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$ and the posterior distributions of τ_1, τ_2, σ^2, ρ_1, ρ_2, and ρ_3 are given in Table 2 when (32) is applied. Meanwhile, when (34) is applied, the MCMC approximation to the predictive distribution of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$ is given in Table 3. By inspecting the two approximations of the predictive distribution of $\mathbf{V}^{(2)}$ given $\mathbf{V}^{(1)}$ and $\mathbf{Y}$, we see that both (32) and (34) yield very similar approximations. The Bayes estimates are computed from the MCMC samples with 50 iterations and 500 replications. Moreover, 30 loops are carried out for each iteration. The convergence of the Gibbs samples is monitored by examining their empirical quantiles.

Table 2. *MCMC approximations to the distributions of* $\mathbf{V}^{(2)}$ *given* $\mathbf{V}^{(1)}$ *and* $\mathbf{Y}$ *and other parameters for the last column of girl data using* (32)

	mean	S.D.	2.5%	5%	25%	50%	75%	95%	97.5%
τ_1	20.233	0.668	18.980	19.122	19.814	20.212	20.671	21.370	21.478
τ_2	0.952	0.158	0.659	0.704	0.847	0.958	1.062	1.221	1.259
$\mathbf{V}_2$	27.731	1.102	25.606	25.959	27.022	27.743	28.468	29.482	29.759
σ^2	4.426	1.732	2.425	2.591	3.316	4.073	5.043	7.464	8.169
ρ_1	0.814	0.080	0.630	0.671	0.773	0.828	0.869	0.917	0.933
ρ_2	0.775	0.095	0.561	0.620	0.725	0.790	0.841	0.896	0.908
ρ_3	0.703	0.129	0.406	0.476	0.624	0.725	0.801	0.868	0.894

Table 3. *MCMC approximation to the distribution of* $\mathbf{V}^{(2)}$ *given* $\mathbf{V}^{(1)}$ *and* $\mathbf{Y}$ *for the last column of girl data using* (34)

	mean	2.5%	5%	25%	50%	75%	95%	97.5%
$\mathbf{V}_2$	27.765	25.476	25.861	26.996	27.759	28.519	29.633	30.005

5. Concluding remarks

The ML and Bayesian methods presented in this paper provide some alternative ways of dealing with the growth curve data when the banded covariance structure holds. The banded covariance structure is definitely one of the more important dependence structure for the general growth curve model in addition to the serial covariance structure.

It is noted that MCMC methods presented in this paper provide superior ways of constructing reliable intervals for the parameters and the future values. Furthermore, the computations involved are relatively easy and should present no difficulty.

Acknowledgements

The authors thank three referees and the editor for carefully reading an earlier version of the paper and for making numerous constructive comments which have improved the presentation of the paper substantially. They are also grateful for research support from the National Science Council of Taiwan (grant NSC 87-2119-M-009-001).

References

Casella, G. & George, E. I. (1992). Explaining the Gibbs sampler. *Amer. Statist.* **46**, 167–174.

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.

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 *Handbook of statistics*, Vol. 1 (ed. P. R. Krishnaiah) 89–115. North-Holland, Amsterdam.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.

Gelfand, A. E., Hills, S. E., Racine-Poon A. & Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Amer. Statist. Assoc.* **85**, 972–985.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. Chapman & Hall, London.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

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

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

Kshirsagar, A. M. & Smith, W. B. (1995). *Growth curves*. Marcel Dekker, New York.

Lee, J. C. (1982). Classification of growth curves, in *Handbook of statistics*, Vol. 2 (eds P. R. Krishnaiah and L. W. Kanal) 121–137. North-Holland, Amsterdam.

Lee, J. C. (1988). Prediction and estimation of growth curve with special covariance structures. *J. Amer. Statist. Assoc.* **83**, 432–440.

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. *Sankyha Ser. A* **37**, 239–256.

Lee, J. C. & Hsu, Y. L. (1997). Bayesian analysis of growth curves with AR(1) dependence. *J. Statist. Plann. Inference* **64**, 205–229.

Lindsey, J. K. (1993). *Models for repeated measurements*. Oxford University Press, Oxford.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.

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. (1987). Prediction of future observations in growth curve models. *Statist. Sci.* **2**, 434–471.

Von Rosen, D. (1991). The growth curve model: a review. *Comm. Statist.* **20**, 2791–2822.

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

Received November 1997, in final form November 1999

Jack C. Lee, Department of Statistics, Columbia University, 2990 Broadway, New York, NY 10027, USA.